

Logical connectives in natural language

A cultural-evolutionary approach

Doctoraalscriptie Algemene Taalwetenschap

Maarten van Wijk

Pelikaanhof 15a, 2312 EA Leiden

06-10386896, mpvwijk@kabelfoon.net

Universiteit Leiden Faculteit der Letteren

Scriptiebegeleider: Dr. P.A. Vogt
University of Edinburgh
Universiteit van Tilburg

Tweede lezer: Dr. C.L.J.M. Cremers
Universiteit Leiden

L.B., september 2006



Abstract

This thesis attempts to explain the particular selection of the 16 logically possible truth-functional connectives that is found in human language. Only a few of these connectives, such as \wedge , \vee and \downarrow (NOR) are expressed using single-morpheme lexicalisations in natural language. Other connectives, such as \mid (NAND) and \leftrightarrow (IFF) are expressed compositionally, by combining words for other connectives or adding extra phrases specific to language.

Various kinds of explanations have been put forward for this observation. *Mentalist-cognitive* explanations appeal to properties and limitations of the logical reasoning mechanisms of human minds (Gazdar and Pullum 1976; Jaspers 2005). On the other hand, *communicative* explanations state that connectives differ in the extent to which they can be pragmatically appropriate or communicatively useful (Gazdar and Pullum 1976; Van Rooij 2005; Sierra-Santibañez 2001).

None of the previous accounts fully answers the research question. Some of the accounts are *functionalistic*: they propose a cause for language's particular set of connectives and stop there (Gazdar and Pullum 1976; Jaspers 2005). What these accounts fail to provide is a clear mechanism that links the proposed causes to language form. The *evolutionary* accounts propose that cultural evolution be that mechanism (Van Rooij 2005; Sierra-Santibañez 2001). However, these accounts use a representation of language structure that is too impoverished to answer the research question.

What is needed is a model of cultural evolution that addresses language surface structure in sufficient detail. Such an approach is the Iterated Learning Model of the emergence of compositionality by Kirby (2000; 2001). This model demonstrates how, as a consequence of the transmission of language across generations of learners, frequent meanings evolve to be represented as irregular holistic phrases, whereas infrequent meanings get compositional representations. The main part of this thesis is devoted to applying this model to the problem of the connectives.

Two routes to a frequency distribution of connectives were pursued. The mentalist-cognitive approach remained unsubstantiated as psychological theories of reasoning difficulty (e.g. Johnson-Laird 2001) were shown to make false predictions about which connectives should be present in language and also unable to provide frequency data in general. More fruitfully, a communicative approach was pursued using a simplified model of human communication. This model, Sierra-Santibañez (2001), consisted of agents aiming to discriminate sets of topic objects from background objects. For this purpose the agents used descriptions involving perceptual properties of the objects, conjoined by the connectives. An exploratory analysis was done of the variables influencing the frequencies of the connectives in this simulation.

Analysis of the simulation results revealed a hierarchy among the 16 connectives with respect to a property of *specificity*, familiar from Gricean pragmatics. It was shown that in a number of situations that are, within the limited bounds of the reality of the simulation, the ones most reminiscent of human communicative situations, \wedge and \vee are the connectives most frequently used by the agents, because they are the ones that conform best to Grice's (1975) Maxim of Quantity. More research is needed on the external validity of the communication model used, however.

A concrete application of Kirby's model proved a bridge too far, as the model simulates the emergence of a different kind of combinatoriality than is found with connectives. Changes need to be made to the learning mechanisms implemented in this model in order to apply it to the case at hand.

Despite the lack of the desired conformation by a computer simulation, connective frequency looks like a strong candidate for an explanation of language's set of single-morpheme logical connectives. In particular, Zipf's (1949) principle of least effort is likely to favour a system in which the most frequently needed connectives are realised as single morphemes from which all others are derived. The communication simulation in this thesis suggests that those most frequent connectives may well be \wedge and \vee or $\underline{\vee}$, because of their usefulness in the communicative situations that humans tend to find themselves in.

Preface

At the end of 2003, three and a half years into my study of linguistics at Leiden, I read an interview with Luc Steels in the Dutch weekly *Intermediair* (Hellemans 2003). The article recounted of communities of robots developing language, *on their own*. ‘They [the robots, MvW] are creating a language we have trouble understanding,’ prof. Steels said suggestively in the article. The sheer futurism of this research got me hooked and led me to discover a branch of my field that I had not known the existence of before: evolutionary linguistics. This is what I wanted to graduate in.

I ended up not writing my thesis at Steels’s Artificial Intelligence Lab in Brussels: instead I went over to the University of Edinburgh for a year. The Language Evolution and Computation Research Unit in Edinburgh is probably the world’s largest community of researchers into the evolution of language. Staying at the LEC was a great experience and has provided with me invaluable knowledge without which this thesis could not have been written.

One of the things I learned at Edinburgh is that you have language evolution simulations and language evolution simulations. The folks in Edinburgh and Brussels each tend to do things their own way. I found one man in Edinburgh trying to close the gap between the two groups: Paul Vogt, who graduated with Luc Steels and also happened to become my supervisor. Like much of Paul’s work, my thesis, too, is indebted to the research traditions of both Edinburgh and Brussels.

Finally, the explanandum of this thesis comes from the field of semantics. In my early years in Leiden, the devoted semanticist Crit Cremers has managed to convince me that meaning itself is the most fundamental and fascinating property of language. My choice to channel the explanatory power of evolution into solving a semantic problem specifically can be traced back to Crit’s influence on my education, even though I failed to comprehend many of the other, finer points of his teachings.

Many others have kindly assisted in the process of writing this thesis. I should name:

Josefina Sierra-Santibañez, for developing a cornerstone simulation of my thesis and sending me an excerpt of its original code, crafted especially for me.

Robert van Rooij, for helping me understand his work on evolutionary signalling games, for his failed attempts at explaining the arcanities of mathematical lattices to me and for buying me an ice cream at EvoLang in Rome.

Jim Hurford, for always knowing the right article to cite in any situation, and pointing me to some of the key articles of this thesis in the process.

Simon Kirby, who introduced me to a simulation of his that spawned the research question for this thesis in its final form.

Jonathan Coe and *Simon Levy*, who helped me while I was trying to construct the pay-off table in appendix 3 using multiple inequality solving in *MatLab*, *Wolfram Mathematica* and *CLIP* for Prolog.

Richard Blythe, who, after I had unsuccessfully struggled with *MatLab*, *Mathematica* and *CLIP* for several days, suggested I should just try and play around with the problem in *Microsoft Excel*. That solved it in half an hour.

Bengt Sigurd, *Katalin Balogne Berces* and many other kind respondents to my queries on *Linguist List*.

Finally, I should thank all my friends in Leiden and Naaldwijk for giving me some respite from the stress I got from writing this thesis, and my parents, without who this thesis would not have been possible at all.

Maarten van Wijk, September 2006

Table of contents

Abstract	2
Preface.....	3
Table of contents	4
1 Introduction	6
1.1 Logic, language and evolution	6
1.2 The structure of this thesis.....	7
2 Truth-functional connectives in logic	8
3 Mapping logical connectives to natural language connectives	9
3.1 Logic as model for the meaning of language	9
3.2 Commonly recognised mappings from logical to natural language connectives	9
3.3 Problems with mapping logical connectives to natural language connectives	11
3.3.1 <i>Different truth functions in natural language than in logic</i>	11
3.3.2 <i>Non-truth-functional overlay for connectives</i>	12
3.3.3 <i>Conjoining other elements than propositions</i>	12
3.3.4 <i>A solution from formal semantics?</i>	14
3.4 Compositional connectives in logic and natural language	15
4 Approaches to explaining natural language's set of logical connectives.....	17
4.1 Mentalist-cognitive, communicative, functionalist and evolutionary explanations	17
4.2 A functionalist, cognitive and pragmatic account for all connectives: Gazdar and Pullum	17
4.3 A functionalist and cognitive account for the absence of NAND: Jaspers.....	19
4.4 The need for an evolutionary explanation: why functionalism will not do	19
4.5 An evolutionary and pragmatic account using Evolutionary Game Theory: Van Rooij	21
4.5.1 <i>Signalling games and Evolutionary Game Theory</i>	21
4.5.2 <i>Introducing connectives into signalling games</i>	22
4.5.3 <i>Is there an Evolutionary Stable Signalling Strategy with all 16 connectives?</i>	23
4.6 An evolutionary and communicative account using Language Games: Sierra-Santibañez.....	24
4.6.1 <i>Language Games</i>	24
4.6.2 <i>Evolving connectives with language games</i>	24
5 The evolution of holistic and compositional connectives in the Iterated Learning Model.....	28
6 Looking for a frequency distribution of logical connectives: the mentalist-cognitive approach	31
6.1 Different processing difficulty may give rise to different connective usage frequencies	31
6.2 Syntactic and semantic models of mental reasoning	31
6.3 Predictions for reasoning errors with the connectives of classical logic.....	32
6.4 Predictions for human preferences for particular connectives	32
6.5 Difficulty and frequency seem unrelated in the case of logical connectives	33
7 Looking for a frequency distribution of logical connectives: the communicative approach.....	35
7.1 The 'use' of connectives: being more specific or allowing more possibilities.....	35
7.2 Quantifying the use of the connectives.....	36
7.3 Repeating Sierra-Santibañez's simulation in order to find a frequency distribution.....	38
7.3.1 <i>Repeating Sierra-Santibañez's simulation using her original settings and algorithm</i>	38
7.3.2 <i>Manipulating Sierra-Santibañez's original settings using a new algorithm</i>	39
7.3.3 <i>Methods of statistical analysis of the data</i>	40
7.3.4 <i>Results from run 1: more meaning space dimensions means easier discrimination</i>	41
7.3.5 <i>Results from run 2: the effects of number of regions, topics and background objects interact</i>	42
7.3.6 <i>Explaining the 'positive' and 'negative' reactions of connectives to N_T and N_B</i>	47
7.3.7 <i>Connective specificity as a function of positivity / negativity and number of DNF disjuncts</i>	51
7.3.8 <i>Relating specificity to connective frequency and finding a satisfactory frequency distribution</i>	52
7.4 Specificity in real world human communication	56

8	The nature of the compositionality of connectives.....	59
9	Conclusions	62
	References.....	64
	Appendices	68
Appendix 1.	The 16 two-placed logical connectives.....	68
Appendix 2.	English language lexicalisations of all 16 connectives.....	70
Appendix 3.	A pay-off table leading to an ESSS with separate signals for all 16 connectives.....	72
Appendix 4.	The effects of N_T , N_B and N_R on N_{disc} for 10 of the 16 connectives.	73

1 Introduction

1.1 Logic, language and evolution

In this thesis, I shall try to tackle a small sub-question of a larger dilemma that has puzzled philosophers since Plato: the different natures of logic and language. The modern discipline of semantics, fathered in the late 19th century by Gottlob Frege and Bertrand Russell, has tried to characterise the meaning of language systematically using various incarnations of formal logic. Although Russell's predicate calculus is still the best system we have for this purpose, the goal of a full formal characterisation has never been achieved. Language always turns out a little bit too complex, too unpredictable, and too irregular to fit perfectly into the ever more intricate moulds that semanticists build it.¹

One domain in which language diverges from logic is that of truth-functional connectives. The formal definition of connectives in the form of truth tables allows for 16 possible two-placed connectives. Yet of only two of them can we confidently state that language uses them: \wedge (AND) and \vee (OR). Some of these connectives, like the connective that yields *true* whatever the truth value of its arguments, remain absent because of their obvious communicative uselessness. Others, such as \mid (NAND) or \leftrightarrow (IFF), appear to be perfectly sound devices for reasoning and communicating: their absence in natural language is harder to explain. Rather than trying to find separate explanations for the presence or absence of particular connectives, this thesis aims to formulate a *general* theory that extends over all connectives and accounts for natural language's particular subset of the 16 possible two-placed truth-functional connectives.

Other than being baffled by this ambitious goal for a master's thesis,² the reader may wonder at this point why anyone would be surprised to find discrepancies between logic and language. Of course language is not like logic! Language is a naturally evolved phenomenon, most likely adapted for communication by human beings. Logic on the other hand is a formal device invented by logicians for all kinds of reasons and purposes, such as curiosity, better reasoning, clearer representation, but certainly not for communication by humans.

Still, the problem we are dealing with here is very real, since the meanings represented by the 16 logical connectives are very real, too. They are not artefacts of the way logic is constructed. In fact, these meanings are expressed in natural language all the time. For instance, the meaning of \mid is contained in *you won't get both ice cream and chips*, a perfectly good English sentence. The fact that none of the world's language uses an equivalent of **you will get ice cream nand chips* cannot be easily explained away as a figment of the logician's imagination.

Of course, we need to look outside of the domain of lonesome old logic in order to explain why language is so different from it. In 1967, Herbert Paul Grice held his William James Lectures in which he proposed that some of the meaning elements of language may be best understood as a product of a principle of communicative co-operation between two human beings. In this thesis, too, viewing language as a tool for communication will be the key to solving the dilemma.

I (and many others) consider it fascinating that we use *and* and *or* instead of *nand* or *iff*. So fascinating, in fact, that I have devoted 70 pages and a year of my life to the subject. Yet, I have encountered quite a few people who felt that the solution was somehow obvious and the question not really worth investigating. \wedge and \vee seem so much more *natural* than \mid or \leftrightarrow , they say. \wedge must be very *fundamental* in representing seeing two things at the same time, they say. \vee is about the *fundamental* concept of two choices, they say.

These are all just assumptions, possibly fed by some sort of cognitive closure we may suffer because

¹ A thorough yet very readable overview of the history of logic, semantics and Western linguistics is Seuren (1998).

² If you are not baffled, consider for instance that Dany Jaspers has recently devoted a 260 page Ph.D. thesis to the universal absence of \mid alone (Jaspers 2005).

we have always used the connectives *and* and *or* in language. Gazdar and Pullum (1976) point out forcefully that the question posed in this thesis is worthy of empirical research:

It is important to see that there are no easy, a priori answers to this question. The connectives discussed in logic textbooks unquestionably owe their familiarity to the prior existence in natural language of truth-functional connectives that wholly or partly correspond to them. Characterizing the properties that determined the natural evolution of the particular natural language truth-functional connectives happens to be an empirical endeavour, and one that seems to us to be of some intrinsic importance and interest. (Gazdar and Pullum 1976: 220).

I wholeheartedly agree. I shall therefore stay away from any *a priori* assumptions, common sense arguments or philosophical debate on the fundamentality of any of the logical connectives. I will treat \wedge and \vee as equals until proven otherwise.

1.2 The structure of this thesis

Chapter 2 is a small crash course in logic, focussing on logical truth-functional connectives. In **chapter 3** we investigate mappings of logical connectives to natural language connectives. Various problems with these mappings will be discussed along with possible solutions.

In **chapter 4**, I devise a taxonomy for the previous accounts given for the set of logical connectives that human languages tend to have. I will conclude that of these accounts, the evolutionary ones are potentially the most convincing, but that they fail to answer the research question of this thesis due to lack of attention to the surface structure of language.

Chapter 5 presents a quick outline of my own evolutionary proposal, aiming to fix the problems with the previous accounts. A model (Kirby 2000; 2001) is introduced that does address language structure in sufficient detail. This model suggests a mechanism through which language may evolve to represent *infrequent* connective meanings as compositional structures, and *frequent* meanings as holistic units. In order to apply the model to the research question we will need (a) a frequency distribution for truth-functional meanings and (b) a meaning representation for logical connectives.

Chapter 6 is an attempt to derive such a frequency distribution from psycholinguistic theories of reasoning difficulty, while **chapter 7** takes a communicative approach: how often do we need a connective? A model of communication is introduced (Sierra-Santibañez 2001) that yields various frequency distributions depending on its parameter settings. An exploratory analysis of the effects of these parameters shows that if settings reminiscent of human communicative situations are used, \wedge and \vee are the connectives most frequently used by communicating agents of the simulation.

In **chapter 8** I look at the other requirement of Kirby's model: a meaning representation for logical connectives. The representation in Kirby's original (2001) simulation is shown to be insufficient. The development of a realistic meaning representation for connectives needs to be deferred to future research. **Chapter 9** contains my conclusions.

2 Truth-functional connectives in logic

In sentential logic, two-placed truth-functional connectives are items that conjoin two propositions. Propositions are statements that can be either true or false in some world. The conjunction of two propositions P and Q forged by a connective C has a truth value that is a function F_C of the truth values of the propositions, where F_C is specified by the connective. This is why these connectives are called *truth-functional*. They are also sometimes simply referred to as logical connectives, contrasting with other, non-truth-functional connectives in natural language such *after* and *because*. Below I present two examples of logical connectives and their functions. In these functions, true is represented as '0' and false is represented as '1'.

$$F_{\wedge}(P,Q): \left[\begin{array}{l} P=1, Q=1 \Rightarrow \wedge(P,Q) = 1 \\ P=1, Q=0 \Rightarrow \wedge(P,Q) = 0 \\ P=0, Q=1 \Rightarrow \wedge(P,Q) = 0 \\ P=0, Q=0 \Rightarrow \wedge(P,Q) = 0 \end{array} \right]$$

$$F_{\vee}(P,Q): \left[\begin{array}{l} P=1, Q=1 \Rightarrow \vee(P,Q) = 1 \\ P=1, Q=0 \Rightarrow \vee(P,Q) = 1 \\ P=0, Q=1 \Rightarrow \vee(P,Q) = 1 \\ P=0, Q=0 \Rightarrow \vee(P,Q) = 0 \end{array} \right]$$

As can be seen from the function definitions above, a connective has an output value specified for each one of four input cases, namely (1) both P and Q are true, (2) both P and Q are false, (3) P is true but Q is false and (4) P is false but Q is true. This output can take on two values: 1 and 0. Therefore, $2^4 = 16$ two placed connectives are logically possible.

In principle, truth-functional connectives could also be defined with more than two arguments, in which case more than 16 different connectives would be possible. Such n -placed connectives are almost never used in classical logic, however, and certainly do not seem to be present in natural language. One-placed connectives, on the other hand, are. Negation (\neg) is the most common one.

Appendix 1 contains a list of all the 16 two-placed connectives. Some of the connectives, such as \wedge and \vee , are well-known and have commonly recognised symbols. Others are less commonly used with less readily recognisable symbols, and again others are almost never used and have no symbols specified. I have invented symbols for the connectives that do not have commonly used ones already. The symbols presented in appendix 1 will be used throughout this thesis.

The truth functions of the connectives are described in appendix 1 by listing all the input cases for which the output value of the connective is true. The four input cases mentioned above are represented as follows for a connective C, assuming a proposition $P \subset Q$:

Input case	Representation in Appendix 1
(1) both P and Q are true	11
(2) both P and Q are false	00
(3) P is true but Q is false	10
(4) P is false although Q is true	01

For instance, the connective \leftrightarrow is true when either both P and Q are true, or both P and Q are false. The truth function of this connective is represented as {11,00}.

3 Mapping logical connectives to natural language connectives

3.1 Logic as model for the meaning of language

It is no coincidence that the names of the connectives \wedge (AND) and \vee (OR) correspond to the existing English words *and* and *or*. Logicians have chosen those names because these particular functions closely mimic the interpretation of *and* and *or* in English: that is, if we make the additional assumption that English sentences are equivalent to propositions and ascribe truth values to them.

For instance, let us assume a set of propositions P, Q, R and S, each of which is equivalent to a natural language sentence.

Proposition	Truth value	Natural language rendition
P	true	[[<i>Brussels is in Belgium</i>]]
Q	true	[[<i>Copenhagen is in Denmark</i>]]
R	false	[[<i>Brussels is in France</i>]]
S	false	[[<i>Copenhagen is in Sweden</i>]]

The conjunctions of these propositions are rendered in English as follows.

<i>Brussels is in Belgium</i> P (true)	<i>and</i> \wedge	<i>Copenhagen is in Denmark</i> Q (true)	= true = true
<i>Brussels is in France</i> R (false)	<i>and</i> \wedge	<i>Copenhagen is in Denmark</i> Q (true)	= false = false
<i>Brussels is in Belgium</i> P (true)	<i>and</i> \wedge	<i>Copenhagen is in Sweden</i> S (false)	= false = false
<i>Brussels is in France</i> R (false)	<i>and</i> \wedge	<i>Copenhagen is in Sweden</i> S (false)	= false = false

The truth values of the natural language fragments equal those of the conjunctions they are renditions of. This process of asserting relationships of equivalence between elements of logic and natural language is known as making a semantic mapping; it is the job of natural language semantics, which uses logic as a model for the meaning of language.

In §3.2, I will look into a number of mappings from logical to natural language connectives that have been put forward. After that, in §3.3, I will show that mapping connectives in logic to connectives in natural language is not always straightforward.

3.2 Commonly recognised mappings from logical to natural language connectives

It is fairly uncontroversial that the well-known languages of Europe and Asia have lexemes representing at least three logical connectives: \wedge , \vee and \neg (e.g. Braine and O'Brien 1998: 51).

However, evidence exists that this is not universally true. Maricopa³, for instance, employs a variety of strategies for expressing \wedge and \vee without a single-morpheme lexicalisation (Gil 1991). One such a strategy is to use bare concatenation, the way English also concatenates AP's in phrases like *a healthy, strong man*, where *healthy*, *strong* is interpreted in the same way as *healthy and strong*. In Maricopa, however, this bare concatenation may also be interpreted as a disjunction, depending on context (Gil's ex. 5a and 6):

³ Maricopa is a Yuman language, spoken by Native Americans in Arizona.

- (1) a. *Johnš Billš v?aawuum*
 John-NOM Bill-ACC 3-come-3-PL-FUT
 ‘John and Bill will come’
- b. *Johnš Billš v?aawuumšaa*
 John-NOM Bill-ACC 3-come-3-PL-FUT-INFER
 ‘John or Bill will come’

The suffix *–šaa* causes the switch to an interpretation with *v*. *–šaa* is normally interpreted as an inferential suffix, both in sentences with and without conjunction (Gil does not provide an example of the latter). Inferential suffixes indicate that the speaker does not have direct evidence for his assertion, like in sentence (1a), but is merely inferring it. The epistemic status of the sentence thus influences the interpretation of bare concatenation as either \wedge or \vee . A similar situation is reported for Upriver Halkomelem⁴, which uses a connective *qə* meaning \wedge or \vee depending on context (Ohori 2004: 57). Specifically, a declarative context allows for a conjunctive reading while an interrogative situation allows for a disjunctive reading.

Other strategies Maricopa employs for representing \wedge involve using the two conjuncts as arguments in a subordinate clause headed by verbs glossed as ‘accompany’ or ‘be together’ (Gil’s ex. 8):

- (2) a. *Johnš Bill uḏaavk*
 John-NOM Bill-ACC 3-accompany-SG-REAL
 ‘John is accompanying Bill’
- b. *Johnš Bill uḏaavk v?aawuum*
 John-NOM Bill-ACC 3-accompany-SG-SS 3-come-3-PL-FUT
 ‘John and Bill will come’
- c. *Johnš Bill uḏaavm ñi?yuuk*
 John-NOM Bill-ACC 3-accompany-SG-DS OBJ-PL-1-see-SG-REAL
 ‘I saw John and Bill.’

Other languages with ambiguous words or constructions for \wedge and \vee include Japanese, Thai and Hua, a language from Papua New Guinea (Ohori 2004). Beja⁵ and Dyirbal⁶ apparently do not have words for \vee either (Gazdar and Pullum 1976).

Claims about other connectives have been made, but most of these have always remained somewhat controversial. The most solid evidence seems to exist for \downarrow : Old English *ne*, or German and Dutch *noch* are likely candidates for this connective which is usually realized with two words, *neither ... nor*⁷, in English (Horn 1989: 256). There are old claims about separate words for \vee and $\underline{\vee}$ in several languages. Latin is said to use *vel* and *aut* for \vee and $\underline{\vee}$ respectively (Quine 1952: 5), Finnish uses *vai* and *tai* and Welsh *neu* and *ynte* (Collinson 1937: 95). On closer inspection of the actual linguistic data these claims seem untenable, however (Horn 1989: 225). It also has been suggested that *if ... then* is a

⁴ Upriver Halkomelem is a Salishan language, spoken by Native Americans around the Fraser River in British Columbia, Canada.

⁵ Beja is an Afro-Asiatic language spoken by about 2 million nomads in Egypt, Sudan and Eritrea.

⁶ Dyirbal is a Parma-Nyungan language, spoken by about 5 Aborigines in North Queensland, Australia.

⁷ Although *neither ... nor* consists of two words and therefore cannot be said to be single-morphemic, the construction smacks of a holistic fixed phrase. The meaning of the whole is not easily deconstructable from the parts. The meanings of *nor* and *neither* are, if anything, equal to the meaning of the construction itself. *Neither ... nor* almost looks like it was formed by a process of reduplication.

lexicalisation of \rightarrow (Horn 1989: 225), but there are some cases where the uses of *if ... then* and \rightarrow diverge significantly (see §3.3.1).⁸ Lastly, some candidates for BUT have been put forward, such as *rather than* (Dieterich and Napoli 1982: 163), and contemporary English *nor* in sentences such as *he is rather poor, nor is he exactly handsome* (Horn 1989: 256).

Finally, the connectives that are *not* present in any language are as interesting as the connectives that are. There is no natural language that has all of the 16 possible two-placed logical connectives. In particular, no known natural language has lexicalized \mid (Zwicky 1973; Jaspers 2005).

3.3 Problems with mapping logical connectives to natural language connectives

3.3.1 Different truth functions in natural language than in logic

The definition of \rightarrow states that if P is false, $P \rightarrow Q$ is always true. This gives rise to some unintuitive interpretations of natural language if we assume that \rightarrow is equivalent to English *if ... then* (Forbes 1994: 49).

Proposition	Truth value	Natural language rendition		
P	false	[[<i>Holland won the 1974 World Cup</i>]]		
Q	false	[[<i>Macintosh is more common than Windows</i>]]		
R	true	[[<i>Paris is the capital of France</i>]]		
<i>If</i>	<i>Holland won the 1974 World Cup</i>	<i>then</i>	<i>Macintosh is more common than Windows</i>	= true???
	P (false)	\rightarrow	Q (false)	= true
<i>If</i>	<i>Holland won the 1974 World Cup</i>	<i>then</i>	<i>Paris is the capital of France</i>	= true???
	P (false)	\rightarrow	R (true)	= true

If \rightarrow were equivalent to *if ... then*, any sentence that starts with *if Holland won the 1974 World Cup* would be true, because Holland did *not* win the World Cup that year. In natural language, however, conditionals are not always true if their antecedents are false. Reasoners rather judge the truth value of such conditionals as indeterminable and take the conditional to be irrelevant (Johnson-Laird and Byrne 1991: 64). One could say they reason with a defective truth table: for the input values 01 and 00, no output is specified.

It is disputed whether English *or* actually corresponds to \vee or to $\underline{\vee}$ (Forbes 1994: 17–18; Horn 1989: 224–225, 394; Gazdar and Pullum 1976: 231). Some authors suggest that it is ambiguous between those two readings (Hurford 1974); the alleged existence of separate words for \vee and $\underline{\vee}$ in some languages would support this, but as noted above, there is no good linguistic data confirming this claim. Others explicitly deny the existence of truth-functional $\underline{\vee}$ in natural language (Barrett and Stenner 1971); any supposed exclusive reading would be due to conversational implicature or some other extra-logical factor. Subjects in psycholinguistic tests do not respond uniformly when tested on how they interpret *or* (Evans and Newstead 1980; Roberge 1978, both cit. in Johnson-Laird, Byrne and Schaeken 1992); they are typically biased towards an inclusive interpretation, but a sizeable minority prefers the exclusive one.

⁸ The lexicalisation *if ... then* consists of two words *if* and *then*. As with *neither ... nor*, I suggest that *if ... then* be treated as a fixed phrase and in this case as a more marked version of *if*, which can be used on its own as well, as in *if the streets get wet, it is raining*. The word *then*, however, cannot be used outside of this construction with *if*: it seems to have no independent meaning.

3.3.2 Non-truth-functional overlay for connectives

And is used quite frequently to indicate a temporal sequence, for instance as in (3) below:

- (3) *Max fell and broke his arm.*

Traditionally the temporal meaning has been taken to be *pragmatically* implied, rather than as part of the meaning of *and* in this case (Schmerling 1975). Others have argued, that temporal sequence is also part of the meaning of *and* proper (Bar-Lev and Palacas 1980). This, however, must lead to the abandonment of the idea that logical \wedge and *and* are equivalents.

The natural language connective *if* implies a causal relationship between antecedent and consequent whereas such a relationship is not implied in logical \rightarrow . Since in sentence (4) below both antecedent and consequent are true, the sentence as a whole would have to be judged true.

- (4) *If Paris is the capital of France then water is H₂O.*

Most human reasoners would judge (4) false though, because the antecedent is completely unrelated to the consequent (Forbes 1994: 83–84). Alternatively, reasoners may assume there must be some causal connection they are perhaps just not aware of.

3.3.3 Conjoining other elements than propositions

The connectives *and* and *or* in English do not just conjoin S's (sentences), but also other syntactic categories such as VP, NP, AP and PP. Examples are presented in (5)–(8) below.

- VP
(5) a. *The football fans sang and danced on the tables.*
b. *This thesis will inspire you or give you a headache.*

- NP
(6) a. *John and Mary walked to school.*
b. *The dog will bite John or Mary.*

- AP (predicative use)
(7) a. *The ball is round and red.*
b. *The ball is red or green.*

- AP (attributive use)
(8) a. *John had to lift a large and heavy box.*
b. *The pork comes with a red or green chilli sauce.*

All these categories, VP, NP and AP, are different syntactic realisations of what are considered to be predicates or object properties in semantics: e.g. the properties of [BEING-SINGING] (5a), [BEING-JOHN] (6) or [ROUNDNESS] (7a). No truth value can be ascribed to predicates: they are of a different type, namely a function that yields a truth value only once applied to an object. This would seem to make the connectives in (5)–(8) non-truth functional.

An early linguistic solution to this problem is to represent the NP and VP examples as syntactic contractions of conjoined S's. Thus (9a) would be derived syntactically from (9b):

- (9) a. *John and Mary walked to school.*
b. *John walked to school and Mary walked to school.*

There are several problems with this approach once NP's not denoting individuals are used (Keenan and Faltz 1985: 4–5). One is the so-called collective reading of NP's. Depending on context, some NP's conjoined by *and* refer to a group of entities rather than to each individual (Link 1998), as in example (10) below:

- (10) a. *John and George met.*
 \neq **John met and George met.*
 b. *John and George carried the washing machine to the kitchen.*
 \neq *John carried the washing machine to the kitchen and George carried the washing machine to the kitchen.*

More problems arise with *quantifiers*, which interact significantly with the meaning of the connectives, as can be seen from (11)–(13) below.

- (11) a. *Some student lives in New Jersey and works in New York.*
 \neq *Some student lives in New Jersey and some student works in New York.*
 b. *Some student lives in New Jersey or works in New York.*
 \neq *Some student lives in New Jersey or some student works in New York.*
- (12) a. *No cat dislikes fish and milk.*
 \neq *No cat dislikes fish and no cat dislikes milk.*
 b. *No cat dislikes fish or milk.*
 \neq *No cat dislikes fish or no cat dislikes milk.*
- (13) a. *All logic theses are boring and lengthy.*
 $=$ *All logic theses are boring and all logic theses are lengthy.*
 b. *All logic theses are boring or lengthy.*
 \neq *All logic theses are boring or all logic theses are lengthy.*

In some cases the syntactic derivation does work, but in some it does not, apparently depending on what quantifier is used. In order to understand this, we need to delve deeply into the intricacies of the set theory, second order predicate logic and generalised quantification, wholly outside the scope of this thesis (for an overview see e.g. Chierchia and McConnell-Ginnet 2000).

The gist of the solution of Keenan and Faltz is to ascribe several homonymic meanings to English *and* and *or* depending on whether they conjoin S, VP or NP, while showing that *and* and *or* have homomorphous denotation functions for those categories.⁹ The denotation function *I* of some Boolean connective *C* is homomorphous for two categories *XP* and *YP* if and only if the denotation of the conjunction by *C* of the intensions of *XP* and *YP* equals the conjunction by *C* of the denotations of *XP* and *YP*:

$$I(XP \ C \ YP) \Leftrightarrow I(XP) \ C \ I(YP). \quad (I \text{ distributes over } C.)$$

This means that applying the 'NP-homonym' at NP-level would yield the same interpretation for a sentence as applying the VP-homonym at the VP-level or the S-homonym at the S-level. That is, for some of the quantifiers: whether *I* distributes over *C* depends on the set-theoretic structure of the denotation of the NP, which in turn depends on whether the quantifier in that NP is what is called upward or downward entailing (cf. e.g. Zwarts 1986: 162–185).

⁹ The denotation function maps the semantic meaning or *intension* of a word to its *denotation* or *extension*, i.e. the objects in the real world it refers to. This distinction was first made by Gottlob Frege, who referred to intension and extension as *Sinn* and *Bedeutung* (see e.g. Chierchia and McConnell-Ginnet 2000: ch. 2.).

An explanation in terms of homonyms appears to be supported by cross-linguistic data: many languages use different connectives for conjoining different grammatical categories. For instance Japanese uses *-to* ‘ \wedge ’ for conjoining NP’s, *-te* or $-\emptyset$ for conjoining AP’s and *-te* or $-\emptyset$ for S’s. Mandarin Chinese has *hé* or *gen* for connecting NP’s, *hé* or *you ... you* for AP’s and \emptyset for S’s (Ohori 2004: 44–45).¹⁰

Finally, another way to approach the problem of seemingly non-truth-functional conjunction is to say that the logical connectives are in fact always set operations. In Possible World semantics (Carnap 1956) the extension of a sentence is not its truth value but the set of worlds in which that sentence is true. It easy to see that if the extension of a sentence S_P is a set of worlds W_P and the extension of S_Q is a set of worlds W_Q , then the extension of $S_P \wedge S_Q$ is $W_P \cap W_Q$. Truth-functional operations can thus be eliminated from the semantics.

From the above discussion, in which phrases referring to propositions are sometimes treated as entities, and sets of entities are conjoined using truth-functional operations, one may rightly conclude that set theory and logic are fundamentally interrelated. Set union and intersection can defined using logical connectives as:

$$\begin{aligned} A \cap B &= \{x \mid x \in A \wedge x \in B\} \\ A \cup B &= \{x \mid x \in A \vee x \in B\} \end{aligned}$$

Even more generally, in the mathematical theory of lattices, union vs. intersection and conjunction vs. disjunction are instances the same lattice operators *join* and *meet*. It just happens that they are joins and meets over the lattices of sets and truth values, respectively (Robert van Rooij p.c.). A number of linguists have argued that the philosophical distinction between truth and reference can be safely abandoned for the semantics of natural language (Carstairs-McCarthy 1999; Hurford In Press).

3.3.4 A solution from formal semantics?

One might naively take the formal semantic discussion above to present an obvious solution to the question of this thesis. If truth-conditionals are eliminated, the problem of the 16 possible connectives disappears: there is only set union and intersection, or lattice join and meet. However, more set operators than these two are possible as well. For instance, nothing would prevent an ‘exclusive union’ set operator $\underline{\cup}$ that would construct, from sets A and B, a new set containing all the elements that are in either A or B, but not in both:

$$A \underline{\cup} B = \{x \mid x \in A \underline{\vee} x \in B\}$$

In fact, set theoretic operations based on each connective could be constructed in this way. *Mutatis mutandis* the same can be done for lattice operations. This just changes our question into why language use the particular set of set-theoretic or lattice operations that it does.

Ultimately, formal semantics cannot provide us with clues to why language lexicalises particular meanings the way it does. The formal apparatus is powerful enough to give us all the 16 possible connectives, and presents no inherent reason why one should be more fundamental than the other.

¹⁰ Secondly, not only do many languages have different renditions of \wedge for conjoining different syntactic categories, such a system is sometimes *not* mirrored in the rendition(s) for \vee in the same language, or vice versa. For instance, Maori uses *me* or \emptyset for \wedge conjoining subject NP, but \emptyset , *aa*, or *hoki* for \wedge conjoining S and VP. For \vee , however, it uses the same morpheme *raanei* throughout all syntactic categories (Ohori 2004: 60).

3.4 Compositional connectives in logic and natural language

English has a way of expressing the meaning of each and every one of the connective meanings. Appendix 2 consists of a list of English language renderings of all the 16 logical connectives. As can be seen from the examples in the appendix, language has several strategies for expressing these meanings.

One strategy is combining connectives, like in logic itself. However, the way connectives are combined in natural language only sometimes mirrors the syntax of classical logic. Some combinations in natural language, such as $\neg(P \vee Q)$, mean different things in logic (and in this case, English) than in some other languages, such as Hungarian (Szabolcsi and Haddican 2004: ex. 1–2):

(14) <i>Mary didn't take hockey or algebra</i> Means <i>Mary didn't take hockey and didn't take algebra</i>	$\neg(P \vee Q)$ $\equiv \neg P \wedge \neg Q \equiv P \downarrow Q$ (following De Morgan)
--	--

(15) <i>Mari nem járt hokira vagy algebra-ra</i> Mary not went hockey-to or algebra-to Means <i>Mary didn't take hockey or didn't take algebra.</i> Cannot mean <i>Mary didn't take hockey and didn't take algebra</i>	$\neg(P \vee Q)$ $\equiv \neg P \vee \neg Q \equiv P \mid Q$ (<i>contra</i> De Morgan)
---	---

The same pattern holds for combinations of *and* and *not* in English and Hungarian: English obeys De Morgan's law $\neg(P \wedge Q) \equiv P \mid Q$, whereas in Hungarian $\neg(P \wedge Q) \equiv P \downarrow Q$. The behaviour of Russian, Serbian, Italian and Japanese is similar to Hungarian, according to Szabolcsi and Haddican.

Other combinations of connectives in logic are just plain ungrammatical in language. For instance, the way $\underline{\vee}$ is composed as $(P \vee Q) \wedge \neg(P \wedge Q)$ is impossible in natural language.

(16) **The tiger came or I brought a spear and not the tiger came and I brought a spear.*

In this case, language uses a different strategy, namely introducing an extra fixed phrase that does not have truth-functional meaning by itself. These fixed phrases vary in the degree to which they are compositional in meaning themselves: they often do contain words that are used as connectives independently and sometimes have a structure mirroring logic. In the case of $\underline{\vee}$, *but not both* is added to *or* 'v' to produce *or ... but not both* ' $\underline{\vee}$ '. We can see how *or ... but not both* mirrors $(P \vee Q) \wedge \neg(P \wedge Q)$: of the two unfamiliar words in this phrase, *but* adopts the meaning of \wedge , and *both* refers to $(P \wedge Q)$. Less transparent is the case of \Leftrightarrow : *and only if* is added to *if* ' \rightarrow ' to produce *if and only if* ' \Leftrightarrow '. In logic, \Leftrightarrow may be derived from \rightarrow as in $P \rightarrow Q \wedge \neg(\neg P \wedge Q)$. In the natural language rendition of \Leftrightarrow , *only if* seems to refer to the exclusion of the condition $\neg P \wedge Q$.

Finally one other natural language strategy for producing the meaning of a connective is to use another connective and reverse the arguments. The meaning ' $P \leftarrow Q$ ' is derived from *if P then Q* ' $P \rightarrow Q$ ' by swapping *P* and *Q*, as in *if Q then P* ' $P \leftarrow Q$ '.

Summarising, language seems to use some of the combinatorial possibilities of connectives in logic, but in some cases it employs strategies of its own to derive connective meanings. Strictly speaking English would not need to use such strategies, though, since the set of connectives it has lexicalised, $\{\wedge, \vee, \neg\}$, is *expressively complete* (cf. e.g. Schumm and Shapiro 1990): all the 16 connective meanings can be derived through some logical combination of the members of $\{\wedge, \vee, \neg\}$. So what stops language from always combining connectives like logic does? One possible beginning of an explanation could be processing difficulties.

Combinations of connectives are compositional in the sense that the meaning of a proposition can be inferred by subsequent application of the functions of the connectives to the outputs of the

connectives they scope over. So, for instance, the meaning of $(P \vee Q) \wedge \neg(P \wedge Q)$ can be inferred to be equal to $P \underline{\vee} Q$ as follows:

\vee		\neg					
P	Q	$P \vee Q$	$P \wedge Q$	$\neg(P \wedge Q)$	$(P \vee Q) \wedge \neg(P \wedge Q)$	\equiv	$P \underline{\vee} Q$
1	1	1	1	0	0		0
0	1	1	0	1	1		1
1	0	1	0	1	1		1
0	0	0	0	1	0		0
\wedge		\wedge		\wedge			

When these derivations are simple, as in the case of BUT, the mechanism of logic is used. Once the derivations get more complex, such as for $\underline{\vee}$ or \Leftrightarrow , language adds specific phrases of its own with no clear analogues in logic. Language's idiosyncratic compositional lexicalisation of connectives could thus be a result of a tendency to avoid nesting functions too deeply. This is however mere speculation; the question at hand is altogether different from the one this thesis attempts to answer.

4 Approaches to explaining natural language's set of logical connectives

4.1 Mentalist-cognitive, communicative, functionalist and evolutionary explanations

The main question that this thesis will try to answer is why of the 16 logically possible connectives only some are realized as single morphemes in natural language, while the rest require paraphrases. The work done into this area so far has been very limited. The explanations given can be divided into two categories: mentalist-cognitive explanations and pragmatic-communicative explanations.

Mentalist-cognitive explanations propose that the representations of logic that humans employ in their minds are different from formal logic itself, or that human brains are for some reason limited in applying the rules of formal logic. For instance, Gazdar and Pullum (1976) propose that humans have difficulty processing negation, and that therefore \neg cannot be a connective in natural language (§4.2). Another example of this approach is Jaspers (2005), discussed in §4.3.

Pragmatic-communicative explanations appeal to the principles of a discipline that was founded with the purpose of explaining the gap between logic and language: pragmatics. These explanations state that connectives differ in the extent to which they can be pragmatically appropriate or communicatively useful. For instance, Gazdar and Pullum believe that a connective such $P \text{ AM } Q$ is communicatively useless: it is logically equivalent to P , so the added effort of using the connective gets no extra information across.

Another dichotomy in the accounts given in the literature is between the *evolutionary* ones and the *functionalist* ones. In principle, both these types of accounts might invoke the same underlying source for some property of language, be it mentalist-cognitive or communicative. A functionalist account stops there, however, while an evolutionary account also provides a causal mechanism through which the underlying source could have influenced the properties of language. The differences between these types of explanations will be considered in more detail in §4.4, after which two evolutionary studies will be discussed in §4.5 and §4.6: Van Rooij (2005) and Sierra-Santibañez (2001). In the following chapter 5, I will then outline what kind of evolutionary approach I think would be most suitable for answering the research question of this thesis.

4.2 A functionalist, cognitive and pragmatic account for all connectives: Gazdar and Pullum

The first and last attempt to provide an exhaustive account for the particular selection of connectives that human languages possess has been provided by Gazdar and Pullum (1976). They formulate three criteria that a logical connective has to meet in order to qualify as a natural language connective.

1. *Compositionality*: A connective C cannot be defined in such a way that one of C 's arguments is superfluous in determining the truth value of the proposition conjoined by C . This means that it must not be the case that $P \text{ C } Q \equiv \neg P \text{ C } Q$ or $P \text{ C } Q \equiv P \text{ C } \neg Q$. The principle of compositionality rules out AM , MA , NM , MN , ALWAYS and NEVER as natural language connectives.
2. *Commutativity*. The linear order in which the two arguments of a connective C are given cannot influence the truth value of the proposition conjoined by C . This means that it must not be the case that $P \text{ C } Q \equiv Q \text{ C } P$. The principle of commutativity rules out BUT , ALTHOUGH , AM , MA , NM , MN , \rightarrow and \leftarrow as natural language connectives.
3. *Confessionality*: A connective C cannot be defined in such a way that a proposition conjoined by C is true when both of the arguments of C are false. This means that it must not be the case that $P \text{ C } Q \equiv 1$ if $P \equiv 0$ and $Q \equiv 0$. The principle of confessionality rules out \downarrow , NM , MN , \leftrightarrow , \rightarrow , \leftarrow , \mid and ALWAYS as natural language connectives.

Together these three criteria rule out all logically possible connectives bar three: \wedge , \vee and $\underline{\vee}$. This is of course a pleasing result, but in order for Gazdar and Pullum's principles to count as a true explanation

for the absence of the other 13 two-placed connectives in natural language, the principles have to be backed up with independent evidence. Otherwise Gazdar and Pullum would be explaining the problem with an appeal to a new formulation of the problem: a case of begging the question.

For the principle of compositionality Gazdar and Pullum give a pragmatic justification. A non-compositional connective would force the language user to utter one conjunct that must be irrelevant to the truth value of the sentence. Thus, the language user would violate Grice's (1975: 46) maxim of *relevance*. This seems like a very credible explanation, also in line with the well known principle of *least effort* (Zipf 1949): why say 'P C Q' (where C is a non-compositional connective) if just saying 'P' conveys exactly the same thing?

Gazdar and Pullum give a mentalist-cognitive explanation for the principle of confessionality. They appeal to psycholinguistic evidence that negations are hard to compute for human minds, and that computation time increases exponentially for each extra negative element added (Hoosain 1973; Clark 1974, both cit. by Gazdar and Pullum 1976). Though this may be true, there is no reason why this computation difficulty would imply that no language could possibly have a non-confessional connective.¹¹ In fact old English *ne* or German *noch* seem likely candidates for a lexicalisation of the non-confessional connective ↓ (Horn 1989: 256).

Gazdar and Pullum acknowledge the existence of *neither ... nor* in modern English, but they accommodate this by proposing that *neither ... nor* is derived syntactically from *either ... or* by incorporation of NEG (e.g. Klima 1964, cit. by Gazdar and Pullum 1976). English *nor* certainly gives the impression of being composed out of *not* and *or*. Etymological sources however suggest it may ultimately stem from a contraction in Middle English of Old English *nawther*, composed of *na* 'no' and *hwaether* 'which of two' (Barnhart 1988). In other languages than English the negative element in the lexicalisation of ↓ is not so obvious either. Swedish, a close relative of English, has *varken ... eller* 'neither ... nor', which does not look at all like *inte* 'not' or like *antingen ... eller* 'either... or' (Bengt Sigurd p.c.). In Hungarian ↓ translates as *sem*, which does not look like *nem* 'not' or like *vagy* 'or' (Katalin Balogne Berces p.c.).

The explanation Gazdar and Pullum give for commutativity falls out of the mentalist-cognitive / communicative dichotomy. Instead it is formal-theoretical: according to Gazdar and Pullum 'underlying structures in language are linearly unordered', a claim which they support with various citations of contemporaneous work in linguistics.

Gazdar and Pullum were working in the now defunct tradition of generative semantics, which means they make 'the assumption that underlying structures in syntax correspond essentially to logical structures' (p. 224). We should note that in 'logical structures' themselves, orderedness abounds. Predicate logic, for instance, consists entirely of predicate-argument relationships, ordered structures *par excellence*. Gazdar and Pullum's hypothesised deep syntactic structure must somehow be able to represent the meaning of these logic structures without replicating their ordered structure. I find this unlikely.

However, if we *do* take the claim of linear unorderedness to be true, such unorderedness does not seem to prevent syntactic languages such as English from having predicates with non-commutative argument structures. Verbs are an obvious example: their arguments, objects and subjects, come in a specific order. Also, language has words that look suspiciously like lexicalisations of non-commutative connectives such as BUT (*rather than* and *nor*, mentioned before in §3.2) and → (*if .. then*). The latter is non-commutative, even if we ignore the parts of the truth table of → that refer to the cases with a false antecedent. A sentence like *if it rains then the streets get wet* is not equivalent to *if the streets get wet then it rains*.

Finally, as a general principle of scientific conduct, I think we first should look for explanations not involving any assumptions about hypothetical underlying structure. This will be my aim in the rest of this thesis.

¹¹ I shall discuss the relationship between psycholinguistic computation difficulty and connectives in more detail in chapter 6. Implications for the principle of confessionality specifically are addressed in §6.5.

4.3 A functionalist and cognitive account for the absence of NAND: Jaspers

The main problem with \mid as a natural language connective is, according to Jaspers (2005), the fact that $P \wedge Q$ presupposes $P \vee Q$. That this is the case can be intuitively felt from the example below (Jasper's ex. 84):

- (17) a. *John is in the garden and Peter is in the garden* entails
b. *John is in the garden or Peter is in the garden.*

Jaspers distinguishes IN-logic and EX-logic. IN-logic is the natural logic of human minds; EX-logic is the formal, consciously developed mechanism of logic. Natural language negation reflects IN-logic. It is unlike logical negation in that it cannot always cancel a presupposition. An example of a presupposition-cancelling negation language is emphatic *not* (Jasper's ex. 367):

- (18) *All gnomes are not married. (Because there are no gnomes!)*

The application of the predicate *married* to *gnomes* presupposes the existence of gnomes. This presupposition is cancelled by *not* here. A negation used a bound morpheme such as *un-* as in (19) may not cancel presuppositions, however. If the set of gnomes is empty, the statement in (19) is not false, like it would be in EX-logic, but meaningless in natural language (Strawson 1950).

- (19) *All gnomes are unmarried.*

By the same reasoning, the presupposition $P \vee Q$ cannot be cancelled by a bound morpheme *n-* 'not' in a fictitious natural language connective *nand* ' \mid ' (Seuren 1985: 260–266). Any sentence containing *nand* would thus be meaningless in natural language logic.

4.4 The need for an evolutionary explanation: why functionalism will not do

Up until now we have seen several explanations for the absence of certain logical connectives in natural language: (1) because these connectives are not communicatively useful; (2) because they are difficult to process for human brains; or perhaps (3) because they do not fit mental logic. These are all functionalist explanations: language has a certain set of connectives, because that is the set that conforms best to the demands of communicative use and to the limitations of human mental abilities. What is missing is an account of the mechanism through which language came to attain this set of connectives that conforms so well to these demands. Disentangling the causal connection between external pressures on language and its form has been identified by Kirby (1999) as the problem of *linkage*. This causal connection is what should be provided by an evolutionary account.

In order to understand the difference between functionalist and evolutionary accounts better, we may draw an analogy with biology. A biological equivalent of a functionalist explanation might be: 'Giraffes have long necks, because it helps them to reach the leaves up in high trees.' Unlike some linguists, biologists have not been content with such kinds of explanations. The missing link in biology is filled in by evolutionary theory, which explains how organisms get to conform to or *adapt* to pressures from the environment. The most important mechanism for adaptation in biology is natural selection: generation after generation, the giraffes with the genotypes leading to the longest necks reproduce in larger numbers, because they are the ones that have access to the best food source.

After the infamous 1866 ban by the *Société de Linguistique de Paris* on papers on the origin of language, it has taken linguists a long time to rediscover the possibilities of applying evolutionary theory to language. Starting in the 1970's there has been a distinct functionalist movement in linguistics which related language properties to external factors such as mental parsing mechanisms (Hawkins 1994) or iconicity (Haiman 1985). This movement and mainstream generative linguistics have remained fairly isolated from each other, though (Newmeyer 1998: ch. 1).

A common argument still used today by linguists against functionalist explanations for linguistic phenomena is that they would lead us to expect such a phenomenon to be present in *all* languages. The explanation would therefore be falsified by just one counterexample in one language or dialect. In the words of John Du Bois:

Volumes of so-called functionalism are filled with ingenious appeals to perception, cognition or other system-external functional domains, which are used to 'explain' why the language simply has to have the grammatical particularity that it does—when a moment's further reflection would show that another well-known language, or even just the next dialect down the road, has a grammatical structure diametrically opposed in the relevant parameter. (Du Bois 1985: 353).

This argument is misguided: the hypothesis that giraffes have long necks so they can reach leaves up in high trees is not falsified once we discover that lions or antelopes from the same area have short necks. The argument does show why a functionalist explanation by itself cannot be enough: we need a causal mechanism that explains why some languages or organisms adapt in a particular way to an outside pressure, and why some do so in other ways.

Only with Pinker and Bloom's (1990) paper 'Natural language and natural selection' interest in the evolution of language has been revived in linguistics: in the decade following this paper the number of articles on language evolution increased tenfold (Christiansen and Kirby 2003: 2–3). Pinker and Bloom's attempt at an evolutionary account of language closely follows the model of biological evolution, and attributes the particular functional properties of language to an innate Language Acquisition Device (LAD), which would have evolved biologically.

If we consider language as a unitary phenomenon, a device for communicating, this proposal seems credible: it is not hard to imagine better communicators reproducing faster. More specific properties of language, however, seem unlikely to cause such differential replication. Such accounts have been suggested by *inter alia* Newmeyer (1991) for several principles of Universal Grammar, such as Subjacency and the Empty Category Principle. Lightfoot's (1991: 69) dryly stated response to Newmeyer sums up the argument against biological evolution of specific language properties: 'The Subjacency Condition has many virtues, but I am not sure that it could have increased the chances of having fruitful sex.' And even *if* the Subjacency Condition had such desirable qualities, there are also neurological limitations on the specificity of the language properties that could emerge through the evolution of innate wiring in our brains (Deacon 1997: 328–334). Most linguists these days agree that if there is an innate LAD, it cannot contain exact specifications of all the details of the human language. It probably just contains *biases* that push language to take on certain forms.

Explaining biases in the brain does not equal explaining the state of human language. We cannot assume a simple mapping of innate preferences onto language features, as has been shown by Kirby, Smith and Brighton (2004). Instead, any bias from an innate LAD is just another factor to which language has to adapt, and as biology has taught us, adaptation can take place in unexpected ways. In short, an appeal to biological evolution does not solve the problem of linkage for language.

The way to solve this problem is by approaching language *itself* as an evolving entity. Language goes through a process of continuous cultural evolution: it changes in non-random ways as it is transmitted between users, and is being learned by successive generations. Cultural evolution is the causal mechanism that could link language form and external pressures. Hurford (1990) has dubbed this the *glossogenetic* approach to language evolution, contrasting with Pinker and Bloom's phylogenetic approach. Through glossogeny, it can be shown that language may acquire certain functional characteristics. Using computer simulations, this has been demonstrated for traits such as compositionality (Kirby 2000), co-existence of regular and irregular forms (Kirby 2001), existence of a fair amount of homonyms but few synonyms in language (Hurford 2003), recursion (Batali 2002) and headedness (Christiansen and Devlin 1997).

I should point out that neither phylogeny nor glossogeny is 'the only' process that gave us the language we use today. Instead both phylogeny and glossogeny play a role in a process of co-

evolution, the former in the evolution of a human brain adapted to learning language, the latter in the adaptation of language to those brains and the communicative needs of its users (e.g. Deacon 1997; Kirby and Hurford 2002). Researchers still differ in opinion on how large the part is that each process plays in explaining human language. Which of these two processes, then, would be the one that could provide the bulk of an explanation for the problem at hand, viz. language's particular set of connectives?

Since it is very unlikely that humans would have an innate specification of what connectives their language can contain, and since any other factor influencing the set of connectives can only do so indirectly through a process of cultural language evolution, the glossogenetic approach will be the viable approach. In pursuing this approach we must set out (1) to find possible biases and pressures that may affect which connectives emerge in language evolution and (2) address the causal process through which these pressures affect language evolution.

In the following paragraphs I will discuss two studies, Van Rooij (2005) and Sierra-Santibañez (2001), that explicitly consider the process through which a system of communication with logical connectives can evolve in response to communicative pressures from the external world—I know of no evolutionary studies that appeal to mentalist-cognitive principles as an external pressure. Both the studies derive the biases in step (1) not from research into the real external world, but instead pre-specify a simple model of the environment that we can modify in certain ways. By experimenting with the properties of the model, insight is sought into the properties of the environment that may influence language's set of connectives.

4.5 An evolutionary and pragmatic account using Evolutionary Game Theory: Van Rooij

4.5.1 Signalling games and Evolutionary Game Theory

David Lewis's (1969) *signalling games* are an application of game theory¹² to linguistic convention. In signalling games, the communicative usefulness of a signal is the factor that decides whether such a signal will be part of an optimal system of communicative conventions.

It is assumed that any signal emitted by a speaker has a certain effect on its hearer, or causes that hearer to perform a certain action. Depending on what the situation is, this effect on the hearer may in turn benefit or damage the speaker. For instance, in a situation in which a juicy apple is hanging from a tree, but unfortunately out of reach for the speaker, uttering 'Could you get me that apple?' will cause the hearer to perform an action that is beneficial to the speaker. In signalling games damage or benefit is pre-specified in utility values or pay-offs that are attached to different combinations of abstract situations and actions.

An optimal signalling strategy is one in which the speaker, for every given situation, produces a signal that will in turn cause the hearer to perform the action with the highest possible utility value to both speaker and hearer.

The question remains how such an optimal signalling strategy would develop in the absence of any conscious design. An answer to that question has been found for regular game theory. In Evolutionary Game Theory (Maynard Smith 1982), successful strategies evolve through a mechanism in which organisms using more successful strategies have a higher chance of reproduction.¹³

Van Rooij (2004) recasts signalling games in terms of Evolutionary Game Theory and proves that given a set of utility values for different situations and actions, a community of speakers will evolve an Evolutionary Stable Signalling Strategy (ESSS) in which a speaker will, for each given situation, produce the message that causes the hearer to perform the action with the highest utility value.

¹² A tutorial in game theory written especially for linguists is Benz, Jäger and Van Rooij (2006).

¹³ Evolutionary Game Theory was originally modelled on biological evolution. It can be applied to cultural evolution as well though, if we do not take the *users* of a communication strategy to be the replicators, but rather the strategies themselves. A strategy with a high pay-off to its users would replicate well by being adopted by many organisms (Van Rooij 2004: note 16).

The notion of an ESSS is somewhat more restricted than that of an optimal strategy in non-evolutionary signalling games. For a situation t_1 to have a separate message in an ESSS, it must be the case that t_1 has an action a_1 associated with it that has the highest pay-off for t_1 *alone*, and a_1 must not be the action with highest pay-off for any other situation t_n . If the latter is the case, the two situations t_1 and t_n call for the same action to be taken by the hearer, and there will be no evolutionary pressure for t_1 and t_n to have different signals corresponding to them. However, such a redundant strategy, using separate signals for situations that call for the same action, could still be optimal in non-evolutionary signalling games.

Since we know what defines an ESSS, and since pairs of situations and messages can be ascribed utility values arbitrarily, for any set of situations a pay-off table with utility values can be constructed in which it is evolutionary beneficial to have a separate message for each situation.

4.5.2 Introducing connectives into signalling games

Conceivably, there could also be situations that are actually conjunctions or negations of situations. Those *complex* situations could have separate utility values for each action attached to them. Let us assume two situations, t_1 [A TIGER IS COMING FOR ME] and t_2 [I HAVE A SPEAR ON ME]. Both t_1 and t_2 could have utterances associated with them that would give evolutionary benefits to the speaker. For instance, in situation t_2 a useful utterance might be ‘So let me have your wife, or I will pierce you’. However, that utterance would not be very beneficial in the case of the complex situation $t_3 = t_1 \wedge \neg t_2 =$ [A TIGER IS COMING FOR ME AND I DO NOT HAVE A SPEAR ON ME]). In this situation, an utterance like ‘Help!’ would probably confer more evolutionary benefit.

There are four connectives that can be described using conjunctions and negations of simple situations: \wedge , BUT, ALTHOUGH and \downarrow . If we assume that $t_1 = [P]$ and $t_2 = [Q]$, then $t_1 \wedge t_2 = [P \wedge Q]$, $t_1 \wedge \neg t_2 = [P \text{ BUT } Q]$, $\neg t_1 \wedge t_2 = [P \text{ ALTHOUGH } Q]$ and $\neg t_1 \wedge \neg t_2 = [P \downarrow Q]$. The simple situations and the complex situations that can be described using \wedge , BUT, ALTHOUGH and \downarrow are the situations that are *actual*. By this I mean that they describe a singular state of affairs that is true or false at a particular point in time.

The situations represented by the other connectives, on the other hand, represent disjunctions of those actual situations and can be used by the speaker to represent his *belief state* about the possibility of several actual situations (Van Rooij 2005)¹⁴. This follows from the insight that each connective can be defined in Disjunctive Normal Form (DNF) as a disjunction of conjunctions.¹⁵ For instance, $P \mid Q$ can be rewritten as $(P \wedge \neg Q) \vee (\neg P \wedge Q) \vee (\neg P \wedge \neg Q) \equiv (P \text{ BUT } Q) \vee (P \text{ ALTHOUGH } Q) \vee (P \downarrow Q)$, while $P \underline{\vee} Q \equiv (P \text{ BUT } Q) \vee (P \text{ ALTHOUGH } Q)$.¹⁶ So, if $t_1 = [P \text{ BUT } Q]$ and $t_2 = [P \text{ ALTHOUGH } Q]$ then $t_1 \vee t_2 = [P \underline{\vee} Q]$. The 16 two-placed connectives can all be represented in DNF with either one, two, three or four disjuncts. Hence these connectives have one, two, three or four situations in which they are true, respectively.

How do we define the pay-off of a particular action with respect to a belief state rather than an actual situation? If we take it that the situation represented by each disjunct in the belief state has an equal chance of occurring, then a situation $t_1 \vee t_2 [P \underline{\vee} Q]$ reflects the belief that there is a 50% chance that $t_1 [P \text{ BUT } Q]$ is the case and a 50% chance that $t_2 [P \text{ ALTHOUGH } Q]$ is the case. So, the pay-off value of some action a_x for $t_1 \vee t_2 [P \underline{\vee} Q]$ is the average of the pay-off of a_x for $t_1 [P \text{ BUT } Q]$ and the pay-off of a_x for $t_2 [P \text{ ALTHOUGH } Q]$.

Now recall from §4.5.1 that a situation t_1 merits a separate message if and only if t_1 has an action a_1

¹⁴ Van Rooij (2005) is an unpublished manuscript; a published summary of this manuscript is Van Rooij (2006).

¹⁵ Disjunctive Normal Form is a standardised or ‘normalised’ way of representing a logical formula. A logical formula needs to be normalised if it is to be used as input to many automated theorem proving algorithms. DNF and its sibling CNF (Conjunctive Normal Form) are thus commonly used in machine learning (see e.g. Chang and Lee 1973: 12–15).

¹⁶ It may seem confusing to represent $P \underline{\vee} Q$ —with exclusive or—as $(P \text{ BUT } Q) \vee (P \text{ ALTHOUGH } Q)$ in DNF—with inclusive or. By definition DNF uses inclusive or, but since $(P \wedge Q)$, $(P \text{ BUT } Q)$, $(P \text{ ALTHOUGH } Q)$ and $(P \downarrow Q)$ are all mutually exclusive, we could replace inclusive or with exclusive or in the DNF-representation without any consequences.

associated with it that has the highest pay-off for t_1 alone, and a_1 is not the action with highest pay-off for any other situation t_n . An example of a pay-off mapping that attributes a unique highest pay-off to $[P \underline{\vee} Q]$ is presented in table 1 below (adapted from Van Rooij 2005: 18). The utility value of an action a_x in a situation t_x is represented as $U(a_x, t_x)$.

Table 1. Example of a pay-off mapping that attributes a unique highest pay-off to $[P \underline{\vee} Q]$. If this mapping characterises the environment, an ESSS will emerge in which there is a separate signal for $[P \underline{\vee} Q]$.			
Situations	Actions		
	a_1	a_2	a_3
$t_1[P \text{ BUT } Q]$	$U(a_1, t_1) = 4$	$P(a_1, t_1) = 0$	$P(a_3, t_1) = 3$
$t_2[P \text{ ALTHOUGH } Q]$	$U(a_1, t_2) = 0$	$P(a_1, t_2) = 4$	$P(a_3, t_2) = 3$
$t_3 = t_1 \vee t_2[P \underline{\vee} Q]$	$AVG(U(a_1, t_1), U(a_1, t_2)) = 2$	$AVG(U(a_2, t_1), U(a_2, t_2)) = 2$	$AVG(U(a_3, t_1), U(a_3, t_2)) = 3$

With this pay-off mapping, a_1 is the best action to perform in situation t_1 , because it yields a pay-off of 4, while a_2 is the best action to perform in situation t_2 , because it yields a pay-off of 4 in that situation. However, if the speaker believes that either t_1 or t_2 may be true (i.e. he is in belief state t_3), performing a_1 will have a 50% percent chance of yielding 4 as a pay-off (if t_1 turns out to be actual case) or 50% chance of yielding 0 (if t_2 turns out to be the actual case). The same is true for a_2 : performing that action means gambling for a pay-off of either 0 or 4, depending on what actual situation turns out to be the case in belief state t_3 . The average pay-off will be 2. So, in belief state t_3 , when either t_1 or t_2 could be true, a_3 is the best action to perform, because a_3 will give a pay-off of 3 no matter whether t_1 or t_2 turns out to be true. If the particular set of situations, actions and pay-off values represented in table 1 exists in the environment, the evolutionary need for $\underline{\vee}$ will arise. This is how far Van Rooij (2005) went in demonstrating the possible scenarios for the evolution of connectives for belief states in signalling games.¹⁷

4.5.3 *Is there an Evolutionary Stable Signalling Strategy with all 16 connectives?*

The crucial question now is whether it will be possible to construct a utility matrix similar to table 1 that models an environment in which there is evolutionary pressure to develop a separate signal for each of the 16 connectives. In such a table, for each connective C there must be a complex situation t_1 conjoined by C , that has associated with it an action a_1 yielding the highest pay-off for t_1 alone and a_1 must not be the action with highest pay-off for any other complex situation t_2 with the same premises but a different connective.

Constructing pay-off tables for actual situations is straightforward, because for those situations the pay-off value of an action can be assigned an arbitrary value. However, for connectives reflecting belief states the pay-offs of each action are calculated as a function of the pay-off values of other actual situations for that action. A change to one pay-off value will affect many other values in the table, and this interrelatedness may make a table with all 16 connectives impossible. If this turns out to be the case, we would have proof from signalling games and Evolutionary Game Theory that there is no evolutionary need for some connectives.

However, after some experimentation I have discovered that pay-off tables yielding communication systems with separate messages for all 16 connectives do exist. Such a table is presented in appendix 3. This finding shows that the meaning of each connective is a potentially useful thing to express. Whether the connective meaning is actually useful depends on the environment, in this case the pay-off table. Importantly, it is not the case that there is no connective that would not be useful in any environment at all.¹⁸

¹⁷ Van Rooij actually constructed $P \underline{\vee} Q$ out of the simple situations P and Q rather than out of $P \text{ BUT } Q$ and $P \text{ ALTHOUGH } Q$, while he had separate explanations for the evolution of \wedge and \neg . I have used the approach with DNF-disjuncts for simplicity, but I believe results similar to mine could be achieved using just simple situations and their negations.

¹⁸ This is even true for the non-compositional connectives, although these are logically equivalent to the simple situations

4.6 An evolutionary and communicative account using Language Games: Sierra-Santibañez

4.6.1 Language Games

Another language evolution simulation paradigm that lends itself to studying the way in which language is shaped by the environment is that of Luc Steels's Language Games, most famously exemplified in his Talking Heads experiment (Steels 1999). In these simulations, robot agents evolve concepts and vocabularies for describing the world around them. Their worlds are filled with simple objects, definable by properties such as their colour, shape or location on a grid. Such properties are dimensions in a *meaning space*, and concepts are defined as intervals on these dimensions. For instance, the concept AT THE TOP might be defined as interval (7,10] on a meaning space dimension with domain [0,10], describing an object's vertical position on a grid.¹⁹

The robots evolve a set of both concepts themselves and names for those concepts by playing series of *Language Games* in which they negotiate word meaning in pairs. Two possible Language Games are the *observation game* (Vogt 2000), in which the robots point out the names of the objects to each other and the *guessing game* (Steels 1999; Vogt 2000), in which one robot makes the other guess which object is meant by a particular word. The guessing game, which we will be concentrating on in this paragraph, proceeds as follows:

- I. A speaker randomly picks an object as a topic and a number of objects as a background.
- II. The speaker tries to find a concept in her conceptual system that uniquely picks out that topic from the background. If none is found, a new concept is created, for instance by dividing an existing concept interval in halves. This part of the guessing game is called the *discrimination game*. The term is somewhat confusing, since the discrimination game is only a subroutine of other language games and unlike those other games it involves only one player.
- III. The speaker lexicalises the concept and communicates the word she has found to the hearer. If the speaker does not have a word for the concept yet, she invents a new word.
- IV. The hearer interprets the word and points to the object she thinks is meant by the speaker.
- V. If correct, the hearer reinforces its conviction that the particular word indeed had the meaning she thought it had.
- VI. If incorrect, the speaker points to the object she had intended, and the hearer responds by adjusting her own lexicon.

Of course in each of these phases there is room for adjusting the properties of the simulation and the way the game is played.

4.6.2 Evolving connectives with language games

This language game model of the emergence of a vocabulary of perceptual concepts has been applied by Sierra-Santibañez (2001) to the evolution of concepts and words for logical connectives. The development occurs in two distinct and consecutive phases.

In the first phase the agents play guessing games in which they develop a lexicon of perceptual categories describing the horizontal and vertical positions of objects on a grid. They also develop a procedure for determining whether some category is true for a particular object or not, and map these combinations of objects and categories to a member of {0,1}. Once this phase is over, the vocabularies and conceptual systems of the robot agents are fixed.

In the second phase the agents evolve concepts for logical connectives by playing *evaluation games*. Evaluation games are like the guessing games discussed in §4.6.1, with a few notable differences. In

¹⁹ t_1 and t_2 . The non-compositional connectives arise because the simple situations they are equivalent to are not present in the pay-off table. It may be presumed that their meanings would usually be expressed in natural language with just the word for the equivalent simple situation, rather than with one of these connectives.

¹⁹ For a more extensive discussion of meaning spaces, see ch. 8.

phase (I), topic selection, a speaker is allowed to pick more than just one object. In phase (II), the discrimination game, the speakers do not just discriminate the topic using a simple predicate, but using a complex expression consisting one or two perceptual categories and a logical category. Such expressions look like simple formulae of propositional logic in Polish notation, i.e. the operator precedes the arguments:

true(up(x))
not(down(x))
iff(up(x), right(x))
and(left(x), up(x))
xor(left(x), right(x))

Note that lexicalisations like *true*, *iff* and *xor* (or *left* and *up*) above are not really used in the simulation; the robots develop their own vocabulary.

In order to find a discriminating expression, a speaker first randomly picks one or two perceptual categories and evaluates them on both the topic and background object(s). This results in two sets of truth values T_{topic} and T_{bg} in the cases where one perceptual category is used, and two sets of truth value pairs $\mathbf{T}_{\text{topic}}$ and \mathbf{T}_{bg} in the case of two categories. If $T_{\text{topic}} \cap T_{\text{bg}} = \emptyset$ or $\mathbf{T}_{\text{topic}} \cap \mathbf{T}_{\text{bg}} = \emptyset$, those categories in fact discriminate the topic from the background. The possible values of T_{topic} and $\mathbf{T}_{\text{topic}}$ represent the connectives. If the speaker does not have this particular set of truth values or truth value pairs as a logical category in his conceptual system, that category is added to the system.

The one-placed logical categories are thus represented as subsets of $T_1 = \{0,1\}$ and the two-placed ones as subsets of $T_2 = \{(1,1),(1,0),(0,1),(0,0)\}$. The latter definition follows from the insight that each connective can be defined in Disjunctive Normal Form as a disjunction of conjunctions. For instance, $P \mid Q$ can be rewritten as $(P \wedge \neg Q) \vee (\neg P \wedge Q) \vee (\neg P \wedge \neg Q)$. This logical category \mid would be represented as $\mathbf{T}_{\mid} = \{(0,1),(1,0),(0,0)\}$ in the simulation: each truth value pair $T \in \mathbf{T}_{\mid}$ corresponds to one disjunct in the DNF representation of \mid and the first and second member of each T correspond to the truth values of the first and second conjunct of each disjunct. As we saw in the §4.5, Van Rooij (2005) used the same normalisation as a basis for his representation of the connectives.²⁰

All this has been rather abstract, so let us move to an example and see how the simulation actually works. Assume that in one evaluation game a speaker needs to discriminate the topic objects TOPIC1, TOPIC2 and TOPIC3 from background objects BG1 and BG2. Let us further assume that the objects have three perceptual features X, Y and Z, each of which can have values ranging [0,10], and several perceptual categories have already been defined over these features. The feature values of the objects are as follows:

	X	Y	Z
TOPIC1	1	4	1
TOPIC2	7	7	7
TOPIC3	9	1	9
BG1	9	1	7
BG2	4	1	1

The speaker always tries discriminating the topic objects from the background by using a single perceptual category and a one-placed connective first, but it can be shown that that procedure would

²⁰ Note the difference between Van Rooij's and Sierra-Santibañez's use of the connectives with more than one DNF-disjunct. Van Rooij uses them to represent disjunctions of possible situations. Sierra-Santibañez uses them to represent disjunctions of objects with differing properties. In either case, these connectives always refer to more than one object or situation.

have failed in this case. I will concentrate on the procedure for discriminating the topic using a two-placed connective.

The speaker randomly picks an ordered set of perceptual categories, e.g. $\{slonk, gnurk\}$. *Slonk* means $Y=[0,3]$ and *gnurk* means $Z=(8,10]$. Let E be the evaluation function over perceptual categories and objects; evaluating *slonk* on TOPIC1 yields $E(slonk, TOPIC1) = 0$, because for TOPIC1 $Y=4$ and $4 \notin [0,3]$. Likewise $E(gnurk, TOPIC1)$ yields 0 as well, leading to an ordered truth value pair $T_{TOPIC1} = \{0,0\}$. In the same way truth value pairs $T_{TOPIC2} = \{0,0\}$, $T_{TOPIC3} = \{1,1\}$, $T_{BG1} = \{1,0\}$ and $T_{BG2} = \{1,0\}$ are derived. $T_{TOPIC} = T_{TOPIC1} \cup T_{TOPIC2} \cup T_{TOPIC3} = \{(0,0), (1,1)\}$ and $T_{BG} = T_{BG1} \cup T_{BG2} = \{(1,0)\}$. The topic is discriminable from the background using these features since $T_{TOPIC} \cap T_{BG} = \emptyset$.

The set T_{topic} turns out to represent the biconditional, but this is not relevant to the computer. If the speaker already has a word representing the logical category matching T_{TOPIC} , say *klok*, meaning $T=\{(0,0), (1,1)\}$, she communicates this to the hearer in a representation $klok(slonk(x), gnurk(x))$ — cf. step III of the guessing game. This representation can be evaluated as true when $T_{klok} = E(slonk(x), gnurk(x))$. It may be paraphrased as: ‘the topic I am trying to communicate to you has $Y \in [0,3]$ if and only if it has $Z \in (8,10]$ ’. The reader may verify from the table that this is indeed a characterisation that picks out the topic objects from the background objects.

Note, that for a particular set of objects several combinations of feature pairs and connectives may exist that discriminate the topic from the background. Only a subset of these will be available to the speaker after the first step of the discrimination game, the random pick of an ordered pair of perceptual features. This step fixes the arguments and their order for the connective that could be used. However, under the limitation of one fixed set of features, too, there can be more than one logical category that discriminates the topic. For instance, in the example above, the topic may also be discriminated with the conditional rather than the biconditional. The conditional would be represented as $T_{\rightarrow} = \{(1,1), (0,1), (0,0)\}$, and since $T_{\rightarrow} \cap T_{BG} = \emptyset$ it also discriminates the topic from the background, in the same way that T_{\leftrightarrow} does. The reason that \rightarrow works as well as \leftrightarrow in this case, is that \rightarrow is true in all the cases that \leftrightarrow is true in. \rightarrow is also true in one extra case (i.e. $\{1,0\}$), but that does not prevent discrimination since the background does not contain any objects for which \rightarrow is true.

In fact, the only way to find a discriminating logical category other than the one picked by the search algorithm is to use a connective with more DNF-disjuncts, as with \leftrightarrow and \rightarrow . Sierra-Santibañez’s algorithm searches for the ‘tightest fit’, however. It simply evaluates the feature pairs on each topic: thus it is guaranteed to find the logical category with smallest possible number of elements. Any other element will be one that is not true of any of the objects in the topic, and thus in a sense superfluous. I will return to this heuristic in §7.3.2 as I discuss whether it is warranted in a model of human communication that aims to be realistic.

After the speaker has communicated her representation to the hearer, the hearer interprets the representation and points to the topic object, as in step (IV) of the guessing game. If the hearer is correct, the next round is played. If she is wrong, the speaker points to the topic, and the hearer essentially repeats the procedure that the speaker has undertaken. First the perceptual categories are evaluated, after which a set of truth value pairs is constructed that discriminates the topic from the background. The hearer then stores this set in her lexicon, together with the word that the speaker had used previously for it.

Since the hearer can already be certain about the fixed meanings of the perceptual categories, and since both hearer and speaker use the same deterministic procedure for finding a discriminatory logical category that, given two perceptual categories, always yields the same result, there is no way that the logical lexicons of the agents could be different at any time. In the guessing game for perceptual categories on the other hand, the hearer cannot always derive the correct meaning of a presented word with certainty from the speaker’s pointing to an object the word refers to. Indeed, the way lexicons align to each other *despite this ambiguity of pointing* is one of the major explananda of Steels’s model. The problem dates back to Quine (1960: 29 ff.) who pointed out that if some native

tribesman exclaims ‘Gavagai!’ while pointing to a rabbit, the linguistic field worker still could not be sure that *gavagai* means RABBIT. It could also mean FURRINESS, FOUR-LEGGEDNESS, RUNNING, etc. In our evaluation game for logical connectives however, we have seen that if a speaker points to a set of objects and utters two known perceptual categories and an unknown connective, the meaning of that connective **can** be unambiguously derived from the pointing, if both hearer and speaker are biased to find the connective with the ‘tightest fit’.

In Sierra-Santibañez’s simulation it turns out that the agents develop all of the 14 of the 16 logically possible connectives during the course of 10,000 evaluation games. The only ones that do not arise are those which yield either false or true in all cases, i.e. ALWAYS and NEVER. Thus all connectives seem to be potentially useful in describing objects. Sierra-Santibañez reports only briefly on the order in which the connectives arise during the simulation and the relative frequencies of their use by the robot agents. She only states that unary categories (TRUE²¹ and \neg) are learnt first, followed by ‘binary connectives such as conjunction, disjunction and implication’, for which lexical coherence is reached after 3,000 games. After 10,000 games all connectives have been learnt.

²¹ TRUE(P) is true if and only if $P \equiv 1$.

5 The evolution of holistic and compositional connectives in the Iterated Learning Model

The evolutionary models considered in chapter 4 have demonstrated that it is useful to have a communication system that can express the meanings of all the 16 connectives. This should not come as a surprise. As can be seen from appendix 2, natural language has evolved to meet this need: it has a way to express the meaning of each connective. However, we still do not know why some truth-functional meanings, such as of \wedge or \vee , are realized as single morphemes and others, such as \mid , as a composition of other connectives. This was the question this thesis was meant to answer.

The models discussed so far concentrate on the evolution of the *meanings* of the connectives. Some representation of language form does co-evolve with the meanings in the models, but that representation is too impoverished to answer the research question. Van Rooij represents the meanings of each utterance with just an abstract symbol t , encompassing both the connective and its arguments. Sierra-Santibañez pre-imposes a structure of two arguments and a single-morpheme connective on the expressions her agents can build.

What we need is a model that addresses the surface form of language in sufficient detail. Specifically, that model should account for the concurrent or sequential evolution of both simple single-morpheme units of meaning and the complex compositional meaning constructs that are the hallmark of modern human language.

A large amount of research has been done into the evolution of compositionality. Two schools of thought exist that divide the evolutionary linguistics community. The first one is the oldest one: the synthetic account envisages language starting its existence in the form of a ‘protolanguage’ with just simple words and no syntax (e.g. Bickerton 1995). Syntax would have evolved later, mainly as an exaptation to other pre-existing cognitive structures of the human brain. For instance, argument structure may have evolved on top of existing mechanisms for determining givers and receivers in social contracts (Bickerton 2000).

The other route to the emergence of compositionality is the analytic one (Wray 1998). According to this theory, protolanguage was a system of *holistic phrases*, similar to animal calls, which functioned as complete messages serving concrete communicative goals. Through various processes of language change, such as re-analysis, these phrases would get broken up into shorter words progressively.

Both these models have their merits. I do not claim to be able to resolve which one is correct.²² In this thesis I shall adopt the analytic approach, though, mainly because this approach has been modelled in interesting computer simulations various times (Kirby 2000; 2001; Vogt 2005; De Beule and Bergen 2006). These simulations do not only confirm that the causal mechanisms proposed in the analytic approach may indeed result in the emergence of compositionality, but also illuminate the factors that determine whether a given meaning will come to be expressed as a holistic or a compositional signal.

In all the simulations mentioned a population of artificial agents is busy trying to get meanings across to each other. The agents have various mechanisms for inventing new concepts, for creating words for those concepts and for inferring what another agent could mean with a certain utterance. New generations of fresh learners are added periodically to these populations, while older speakers ‘die’ and are removed from the simulation. Because language is learned over and over again by successive generations in this kind of models, they have been dubbed *Iterated Learning Models* (ILM’s) by Kirby and colleagues (e.g. Kirby and Hurford 2002). I shall not go into how these simulations work precisely, and concentrate instead on what factors in each model determine whether

²² The incessant and frequent debates between Jim Hurford and Simon Kirby on this topic have become somewhat legendary at the LEC in Edinburgh. For years now, the two of them have been hijacking countless discussion sessions meant to be on other topics, topics often only vaguely related to the issue of the origins of compositionality.

a given meaning will come to be expressed as a holistic or a compositional signal.

In one important ILM simulation by Kirby (2000) the factor causing compositionality in language is a *bottleneck effect* in its transmission across generations (Kirby 2002: 20–21). Language learners are always presented with just a subset of the language they are trying to learn: this is the bottleneck, through which elements of a language must pass again and again as it is learned by a next generation. Compositional rules are more likely to pass through the bottleneck than holistic ones, since they can be used for the lexicalisation of more than one meaning. Hence they are more frequent, and are thus more likely to be presented to the next generation somewhere during their learning period.²³

However, holistic rules can stand a good chance of passing through the bottleneck as well if their meanings need to be lexicalised frequently enough. If a Zipfian frequency distribution²⁴ is introduced over the meanings in a computer simulation, it can be shown that after many generations of learners, the most frequent items get realised as irregular, holistic signals, whereas less frequent meanings are realised regularly and compositionally (Kirby 2001).

In another model of the emergence of compositionality by Vogt (2005), some other factors promote the emergence of compositional rules. Importantly, in Vogt's model too, frequency helps a rule to survive, but not just because it makes the rule more likely to pass through the bottleneck.²⁵ In Vogt's model, every time a rule is used successfully in an interaction, its weights are reinforced making it more likely to be used in future, whereas other rules competing for the same meaning are inhibited. Vogt does not provide an account of regulars versus irregulars, but given the role of frequency in this model it may be assumed that frequent meanings would be lexicalized irregularly in this model, too (Paul Vogt, p.c.). The same basic mechanism of reinforcement of successful items is used in a different study by De Beule and Bergen (2006). If a frequency distribution were introduced here, frequent holistic phrases would tend to survive as well (Joachim de Beule, p.c.).

Frequency as an explanatory factor for language properties is no stranger to the field of linguistics. Difference in frequency has been invoked to explain *inter alia* the fact that the 10 most frequent English verbs are all irregular (Pinker 1999: 123–125), grammaticalisation (Bybee 2003), relative word length (Zipf 1935)²⁶, markedness, the loss of the OV word order in Old English and differential diffusion of sound changes through the lexicon (all three mentioned in Newmeyer 1998: 123–126).

By picking these models we have taken an important step towards an evolutionary explanation for language's set of natural connectives. Recall from §4.4 that a proper evolutionary explanation would consist of (1) possible pressures or biases that influence languages properties and (2) an account of the causal process through which these pressures influence those languages properties. The computer model provides the causal mechanism, and also a mediating factor through which external pressures may influence the lexicalisations of the connectives: frequency. We are now left with step (1), finding the pressure that could lead to an unequal frequency distribution of the connectives. In particular, we would need proof showing that, for some reason, language users will express some connective meanings, such as \wedge and \vee , more frequently than others, such as $|$ or \leftrightarrow .

In chapters 6 and 7 I shall investigate the two main lines of explanation (mentalist-cognitive and communicative) that have been invoked in the literature to explain connective use in human language and pursue them to see if they could give rise to a frequency distribution.

²³ Van Rooij suggests that the ILM might provide a way in which presumed holistic signals for complex signalling game situations could be broken down into compositional ones containing separate signals for the connective and its two arguments. Note that this would be compositionality arising in a different meaning space than the compositionality this thesis is looking into, namely the breaking up of connectives themselves into meaning components.

²⁴ A distribution of meanings is Zipfian if the relative frequency of each meaning is inversely proportional to its frequency rank (Zipf 1935). Corpus research shows that the natural languages tend to conform to Zipf's law quite well (Li 1992).

²⁵ Bottlenecks can be introduced into Vogt's model, but do not have to in order for compositionality to emerge.

²⁶ Zipf observed that word length is inversely related to frequency without quantifying this relation. A formula has been devised later, though: it turns out to vary according to what language corpus is tested (Sigurd, Eeg-Olofson and Van de Weijer 2004). The observation of Zipf about frequency and word length should not be confused with his law about frequency and frequency rank, discussed earlier in this paragraph.

In chapter 8 then I seek to fulfil another requirement of these models. All the models are about agents learning to convey meanings to each other through signals, and simply pre-specify the meanings. I too shall make the assumption that language users had the meaning of all 16 connectives available before they actually started developing words for them²⁷. These meanings need to be represented in some way. Kirby's (2001) study into the emergence of compositionality and irregulars uses very simple meaning representations: strings of 2 characters from a 5-letter alphabet. Vogt (2005) actually co-evolves meaning with form, but the variation of meanings that may evolve is limited to different variants of bit-vectors. Other studies, such as Kirby (2000) and De Beule and Bergen (2006) use a form of predicate logic. In chapter 8 I shall investigate what kind of meaning representation would be suitable for the meanings of the connectives.

²⁷ How connective meanings (rather than words) may evolve is demonstrated in the models of Sierra-Santibañez and Van Rooij (§4.5 and §4.6). Sierra-Santibañez's model, specifically, co-evolves meanings with words for those meanings. Her words however are always random mono-morphemic strings, while the meanings of the connectives are simply invented the first time a connective is needed for discrimination. A tighter integration between evolution of meaning and words still has intuitive appeal, and it may perhaps be achieved in Vogt's (2005) model that integrates elements of the ILM and Language Games. The matter of the relative timing of the evolution of words for and meanings of the connectives remains an empirical question: there is no *a priori* reason to reject my assumption that connectives meanings came first, and words second.

6 Looking for a frequency distribution of logical connectives: the mentalist-cognitive approach

6.1 Different processing difficulty may give rise to different connective usage frequencies

There is a large body of psycholinguistic evidence for differences in processing difficulties with the various connectives. From the relative difficulty of the connectives we may be able to infer different frequencies of use: one would assume that a connective that is difficult to understand or reason with would tend to be used less often.

One limitation here is that processing difficulty data only exists for a limited number of connectives, namely \wedge , \vee and \rightarrow . The reasons for this are obvious: these are the most common connectives by far in logic and natural language. Research into other connectives is sparse to non-existent. However, the theories advanced in psycholinguistics are general enough to make predictions about the difficulty of the connectives other than those that have been experimented with.

6.2 Syntactic and semantic models of mental reasoning

Two main models have been proposed in the psychology of deductive reasoning: a syntactic model and a semantic model. Both are inspired by inference mechanisms of formal logic: the syntactic model by proof-theoretic deduction, the semantic model by model-theoretic deduction.²⁸

The syntactic model (e.g. Rips 1994) states that humans always automatically abstract the ‘surface form’ of an argument into a logical form. When reasoning they mentally apply a subset of the rules of proof-theoretic inference. Among these are *disjunction elimination* ($P \vee Q$, $\neg Q$, therefore P) and *modus ponens* ($P \rightarrow Q$, P , therefore Q). However, some other rules are usually not included in the hypothesized mental rule system, such as *modus tollens* ($P \rightarrow Q$, $\neg Q$, therefore $\neg P$).

Model-theoretic inference works by setting up models of states of affairs based on the semantics of the premises. In formal logic the semantics or meanings of propositions are represented as truth tables. These truth tables are highly redundant as representations, however. In psycholinguistic models therefore they are compressed in order to reduce the load on working memory (Johnson-Laird, Byrne and Schaeken 1992). Reasoners are taken to represent only the models that are true for a particular connective, and by default only the true propositions in those models (Johnson-Laird 1999: 116). For instance, the proposition $P \rightarrow Q$ would have the following truth table:

P	Q	$P \rightarrow Q$
1	1	1
1	0	0
0	1	1
0	0	1

Compare the models that $P \rightarrow Q$ would be represented by:

1. P Q
2. Q

Each of the models represents a case in which $P \rightarrow Q$ is true. The fact that P is false in model (2) is not represented. Finally, the semantic model states that initially, reasoners do not fully flesh out all of

²⁸ For discussion of proof-theoretic and model-theoretic induction in logic, see e.g. Forbes (1994). For an overview of psycholinguistic research into deduction, including a comparison of the syntactic and model-theoretic models, see Johnson-Laird (1999).

these possible models. Instead, they start with just one model, representing the most salient case. In the case of $P \rightarrow Q$, this is the first model in which P and Q are both true.

6.3 Predictions for reasoning errors with the connectives of classical logic

The two models make different predictions about the errors humans make while reasoning.

The syntactic model predicts that the number of inferential steps needed correlates with the number of errors. However, this number of inferential steps needed depends on the mental rules that are hypothesized. For instance, take the modus tollens inference. If there were a mental rule for this inference it should be easy to make, as easy as, say, disjunction elimination. In fact, reasoners consistently find modus tollens difficult to make.²⁹ Rule theorists explain this by hypothesising that there is no mental rule for modus tollens, and that reasoners need a multiple-step inference, involving modus ponens and a *reductio ad absurdum*. This argument seems circular: reasoners have difficulty with modus tollens, so they must lack a mental rule for it; and because reasoners lack this mental rule, they have difficulty with the inference.

The semantic model predicts that the more models are needed to make an inference, the more errors will be made, because people try to minimize working load by focussing on just one possible model. Evidence for this tendency has been found in spatial, temporal and quantified reasoning as well as logical inference, the topic at hand (Johnson-Laird 2001: 435; Sloutsky and Goldvarg 2004: 638). The difficulty reasoners have with modus tollens is explained as follows.

When presented with the proposition $P \rightarrow Q$, reasoners initially represent it with only one model, and include a footnote (represented by ‘...’ below) that more models are possible:

1. P Q
2. ...

From this initial model and $\neg Q$ nothing follows. This is in fact the response of many untrained reasoners who are invited to make a modus tollens inference. In order to make the inference, the representation for $P \rightarrow Q$ needs to be fully fleshed out, as in:

1. P Q
2. Q
- 3.

Now it can be seen that there exists no model in which $\neg Q$ is true and P is true as well. Thus $\neg P$ follows. So, reasoners need to consider all three models under which $P \rightarrow Q$ is true, and that is what makes modus tollens hard. Note that in order to make the easy modus ponens inference reasoners only need to consider the first model that is initially represented.

6.4 Predictions for human preferences for particular connectives

Whichever theory is correct, what concerns us here is which predictions they make for a preference for one connective over another. These predictions should be extendible to other connectives than the ones of classical logic.

The syntactic theory makes no such prediction (Johnson-Laird 1999; Rader and Sloutsky 2001). Without modifications, it also does not allow for connectives other than the ones already present in classical logic, since the mental rules it proposes are designed to reason with those connectives only.

²⁹ Note that by ‘reasoners’ I mean participants on reasoning experiments with no particular experience in formal logic. Of course anyone can be trained in formal logic, and those that are may become adept at it to such an extent that modus tollens no longer forms a problem. However, the research described here is meant to investigate the everyday reasoning skills human naturally use in their daily lives, not the formal skills of logicians.

A rule-based theory would thus give an a priori account for the particular sets of connectives there are in human language: we have the connectives we have, because they are the only ones we can reason with using our mental rules. From other connectives no inferences can be made at all. They are thus meaningless, and human language has no need for words without meaning.

On the other hand, syntactic theory does not offer any reason why humans would use the inference rules of classical logic, rather than some other arbitrary set of valid rules. For instance, a valid rule for making deductions with \leftrightarrow would be: $P \leftrightarrow Q$; P ; therefore Q . The theory might also introduce a derivational system that rewrites the other connectives into the classical ones. Thus, the model would in fact be compatible with any set of connectives, without giving us any reason why one would be used rather than another.

Semantic theory does have something to say about the individual connectives. If inferences requiring more than one model are harder to make, then propositions containing connectives compatible with more than one possible model will, on average, cause more representational errors (Rader and Sloutsky 2001: 839; Johnson-Laird 2001: 435). However, the number of models a proposition is compatible with does not map one-to-one onto the number of cases in which the connective is true. Johnson-Laird, Byrne and Schaeken (1992) found that the order of difficulty of the connectives is, from easy to hard, \wedge , \rightarrow , \vee and \perp . Note that even though \rightarrow is true in more cases than \vee is (3 versus 2), people find it easier to reason with. This would be the case because reasoners *initially* represent \rightarrow with only one model. As we saw, this limited initial representation makes modus ponens inferences easy. On the other hand, however, it inhibits the modus tollens inference, for which reasoners need to flesh out the initial representation of \rightarrow to include more models. So, the difficulty of a connective depends on the inferences we want to make with it, because different inferences require different amounts of explication of the models with which a connective is compatible.

It will be hard to predict what kind of inferences would be made with hypothetical connectives; in fact we do not even know what inferences people most commonly make with the existing connectives. Also, it is not clear what the model theory would predict about these inferences, because there seem to be no general laws governing the pick of an initial model for a connectives. The theory proposes that reasoners initially postulate a model ' P Q ' for \rightarrow , and two models for \vee , namely ' P ' and ' Q ', but it is not clear why this would be the case. We have no way of telling with what initial representations a reasoner would represent a non-classical connective. Since these are supposed to predict the difficulty of an inference, we have no way of knowing which connectives should be difficult and which connectives easy.

There is a way around this problem if we just consider the case of the connectives with only one condition of truth. These could be equally difficult as other connectives with more conditions of truth, because those connectives might have only of those conditions fleshed out. However, they could never be *more* difficult. If ease of reasoning were a large factor in the evolution of connectives, we would predict at least some languages with the connectives \downarrow , BUT and ALTHOUGH. However, we find none with BUT and ALTHOUGH, and all languages have \vee , which is supposed to be much harder to reason with according to the model.

6.5 Difficulty and frequency seem unrelated in the case of logical connectives

Psycholinguistic theories of reasoning difficulty seem to have little predictive value of the particular connectives that languages contain. Natural language may contain connectives that are notoriously hard to reason with, whereas other connectives, that should be very easy according to some theories, remain absent. It is also far from clear how a frequency distribution for connectives could be derived from processing difficulty differences. However, if somehow a function could be found that predicts higher frequencies for theoretically more difficult connectives, it is obvious that such a function would make false predictions as well.

Our discussion of processing difficulty does shed some new light on Gazdar and Pullums's principle of confessionality, which appeals to the difficulty of the processing of negation in order to rule out connectives which yield true for two false arguments. The difficulty with the processing of | found by Johnson-Laird, Byrne and Schaeken (1992) confirms this, although they cite a different cause: | requires three mental models. On the other hand, semantic theory predicts that another non-confessional connective, ↓, should be easy. In fact, ↓ is found in many languages, and | is not. This should of course not be taken as evidence that the model theory predicts the presence of certain connectives in language, but merely that Gazdar and Pullum's theory about non-confessionality must be mistaken for two reasons: (1) the class of non-confessional connectives is not homogeneous with respect to difficulty and (2) difficulty is not a sufficient condition for the absence of a particular connective in human language, which in fact is shown by the presence of ↓ in Old English, German and Dutch mentioned in §3.2.

7 Looking for a frequency distribution of logical connectives: the communicative approach

7.1 The ‘use’ of connectives: being more specific or allowing more possibilities

All in all processing difficulties do not seem to give promising clues to the origins of languages’ particular set of connectives. In this chapter I shall look into another possible factor: the communicative use of a connective’s meaning. Could it be the case that \wedge and \vee are somehow more useful than other connectives, so that their meanings are needed more frequently? And could this in turn cause \wedge and \vee to be quickly and efficiently realised as holistic phrases, whereas other connectives do with paraphrases?

It is going to be difficult to theorise about which connective meaning would be ‘needed’ how often.³⁰ A good place to start would be an informal description of when and why humans use *and* and *or*, the existing connectives in natural language. As noted in §3.3.3, these connectives can conjoin a number of syntactic categories, such as S, NP and VP. In (20)–(24) below I present some examples of different syntactic categories being conjoined by *and* and *or*, repeating the examples given in (5)–(8) in §3.3.3. For each example, I briefly describe the purpose the connective is used for in that example.

- S
- (20) a. Conjoined by *and*: to describe two simultaneously occurring situations
The box is wet and the vase is wet.
- b. Conjoined by *or*: to describe two possible situations
The box is wet or the vase is wet.
- VP
- (21) a. Conjoined by *and*: to describe a set of actions
The football fans sang and danced on the tables.
- b. Conjoined by *or*: to describe two possible actions
This thesis will inspire you or give you a headache.
- NP
- (22) a. Conjoined by *and*: to describe a set of two persons or objects
John and Mary walked to school.
- b. Conjoined by *or*: to describe two possible persons or objects
The dog will bite John or Mary.
- AP (predicative use)
- (23) a. Conjoined by *and*: to describe two properties of one object
The ball is round and red.
- b. Conjoined by *or*: to describe two possible properties that an object could have
The ball is red or green.

³⁰ At first sight it may seem that no connective is strictly necessary, since the meaning of each connective can be always be expressed in terms of certain subsets of other connectives. A subset of connectives that can express all other connectives through some combination of its members is called expressively complete: such sets are *inter alia* $\{\vee, \neg\}$, $\{\wedge, \neg\}$, $\{\downarrow\}$ and $\{\downarrow, \neg\}$ (Schumm and Shapiro 1990; Zyliński 1924; Sheffer 1913; Peirce 1933). If we have any of these sets, we might say the other connectives are not useful at all. However we should bear in mind here that we are looking for a frequency distribution of truth-functional meanings rather than a distribution of connectives per se.

AP (attributive use)

- (24) a. Conjoined by *and*: to describe two properties of one object
*John had to lift a **large and heavy** box.*
b. Conjoined by *or*: to describe two possible properties that an object could have
*The pork comes with a **red or green** chilli sauce.*

A pattern leaps out from the examples: *and* is used for referring to two properties, objects, actions, etc. simultaneously, whereas *or* is about describing two different states that those properties, objects and actions could take on.³¹ *And* is about being more specific about one particular entity or action; *or* is about admitting the possibility of more entities or actions, that is being *less* specific.

As we noted before when discussing the approaches by Van Rooij and Sierra-Santibañez to accounting for the connectives in human language (§4.5 and §4.6), connectives can be described using Disjunctive Normal Form and split up into categories that yield true in either 1, 2, 3 or 4 states of affairs.

And is in the category that is only true in one case. For instance, example (22a), *John and Mary walked to school* is true if and only if both John and Mary walked to school. The fact that \wedge is two-placed makes it suitable for describing two situations being true simultaneously, namely John walking to school and Mary walking to school. Other connectives in this category are BUT, ALTHOUGH and \downarrow .

Or is in the category that is true in several cases, namely three. For instance, example (22b), *John or Mary walked to school* is true if and only if one of these three conditions is met: (1) John walked to school but Mary did not, (2) John did not walk to school although Mary did, or (3) both John and Mary walked to school. Other connectives in this category are \leftrightarrow , \vee , \rightarrow and \mid .

So, connectives can be split up into two groups depending on the direction in which they diverge in **specificity** from a baseline expression P. \wedge , BUT and \downarrow introduce more restricted truth conditions by adding a conjunct, and thus reduce the number of worlds in which P is true. The other two-placed connectives introduce more lax truth conditions through the different DNF-disjuncts and thus increase the number of worlds in which P is true. Increased specificity can be used to reduce the number of objects, situations, properties being referred to; decreased specificity can be used to increase that number. The question remains: how frequently do we use either more specific and less specific connectives?

7.2 Quantifying the use of the connectives

We saw in §7.1 that some connectives serve to increase specificity, whereas others decrease specificity by allowing more possible situations. Both functions could be useful to a human communicator at some point: he may need to be specific in some cases, and need to make broad statements in other cases. It seems pretty much impossible to quantify these communicative needs. How many times does a communicator need to be specific? How many times does he need to be general? The answer would depend on the *environment*, the *goals* of the communicator and the properties of the rest of the communicative system. It is obvious that these factors are extremely complex. The chances are slim for us to be able to derive any frequency data from them about the need for the 16 logical connectives. Yet *real* frequency data is what we need if we want to use the Iterated Learning Model to find out why some connectives are realised holistically and some are not.

As is common in science when problems become too complex to solve, I shall attempt to reduce the problem to a model. Frequency data may be obtained from some simplified model of communication. Two of such simplified models of communication have already been discussed in this thesis: Van Rooij (2005) in §4.5 and Sierra-Santibañez (2001) in §4.6.

³¹ Admittedly, this fact is not exactly ‘leaping out’ of raw data, but rather out of my interpretation of the data, which may be biased. The reader is invited to verify my interpretation and judge whether it makes any sense.

In Van Rooij's model, a *communication system* is evolved through the extra-linguistic mechanisms of Evolutionary Game Theory. After evolution, we find that this communication system consists of a simple one-to-one mapping of abstract situations and signals. Some of these signals contain the meaning of one of the 16 connectives. At what point such a signal containing a connective is produced depends on what abstract situation is the case. This in turn is determined entirely arbitrarily. There is no way we could obtain any meaningful connective frequencies from this model.

In Sierra-Santibañez's model the *environment* consists of a set of objects. The *goal* of the speaker is to point out a set of topic objects to the hearer, in such a way that the hearer cannot confuse them with a different set of background objects. It is assumed that the particular set of topics and background objects the speakers needs to point out changes randomly on each occasion. Every such occasion is modelled in one evaluation game. The *communication system* consists of expressions for simple perceptual properties of the objects, which can be used as arguments to one and two-placed connectives.

If we want to understand what part of real human communication is being modelled here, we could imagine two children (*speaker* and *hearer*) sitting in a room full of toys (the *objects*). At some point (one *evaluation game*), one of the children wants to have two particular toys, a green one and a red one (the *topic* objects). Those toys lie in one corner of the room along with a few other toys, say a pink one and a blue one (the *background* objects). The child wants his companion to get him the green toy and the red toy. Therefore she asks: '[Could you get me the toys that are] *red or green*?' because red or green is the qualification that is true of the objects the speaker wants to have and false of the other toys. If the other child then hands the speaker the desired toys the speaker has achieved her goal.

In short, the linguistic behaviour modelled is that of humans pointing out objects to each other by describing the properties of those objects. It most closely resembles the use of connectives in (23)–(24): to conjoin AP. This may not seem very 'useful' or common behaviour, but according to one prominent evolutionary anthropologist and cognitive scientist (Tomasello 1999), the habit of calling other people's attention to objects and establishing *joint attention* is specifically human behaviour that is fundamental to the development of our cognition.

Although Sierra-Santibañez makes no mention of frequencies herself, her simulation could potentially provide us with frequency data about the use of all the connectives. Over the course of a number of discrimination games, we can count the number of times a speaker could use a particular connective to build a discriminating expression. This means that we shall use Sierra-Santibañez's original simulation for a slightly different purpose than it was set up for. Sierra-Santibañez was researching how concepts and words for connectives could emerge in some community of artificial speakers and hearers, as a consequence of a need to point out sets of objects. I am looking for how often the meaning of each connective is useful to speakers for pointing out the desired object, assuming the meanings of the connectives are already available to the speaker. I also aim to make this frequency distribution explain properties of the real world, so unlike Sierra-Santibañez I want the speakers and environment to be as realistic as possible.

In the following paragraphs I will use Sierra-Santibañez's simulation to obtain a frequency distribution of logical connectives. It will turn out in §7.3.1 that Sierra-Santibañez's original algorithm yields a frequency distribution that would produce unrealistic lexicalisations of the connectives if used as input to Kirby's model.³² These results can partially be ascribed to some unrealistic parameters of the original algorithm.

These parameters are discussed in §7.3.2; they include properties of the speaker's cognitive system, the structure of the environment and the communicative system. Appropriate settings for some of these parameters are suggested, so that the simulation can be made more realistic. For some other

³² It is easy to predict which meanings will be realised holistically and which ones compositionally in Kirby's (2001) model: it is always the most frequent ones that get lexicalised as holophrases. An unrealistic frequency distribution is therefore a distribution in which \wedge and \vee are *not* the most frequent connectives.

parameters, it is proposed that the effect should be researched. In order to make these changes and to be able to tweak the parameters easily, I explain how I have reimplemented the simulation in a different programming language.

Then, in order to find out what exactly is the effect of these parameters, §7.3.3 proposes to run a large number of simulations, in which different combinations of parameter settings are tested. Using statistical methods, then, the effects of the different properties can be determined.

The result of one preliminary run of simulations is presented in §7.3.4. The results from this run prompted a second run of simulations. The results of this second run are presented in §7.3.5. The main observation we make from the data is that the connectives can be divided into two groups, positive and negative, that respond differently to different numbers of topic and background objects. The emergence of these groups is analysed and explained formally in §7.3.6. In §7.3.7 I then relate these groups to the notion of specificity already touched upon in §7.1. Finally, in §7.3.8, I discuss how specificity affects the frequency of use of the connectives.

7.3 Repeating Sierra-Santibañez's simulation in order to find a frequency distribution

7.3.1 Repeating Sierra-Santibañez's simulation using her original settings and algorithm

In the process of finding a frequency distribution of connectives, I have initially used the same Prolog discrimination game algorithm as used by Sierra-Santibañez for her (2001) article. I have limited myself to the discrimination game part, and have ignored the part where the hearer interprets the speaker's utterance. This is because I aim to use Sierra-Santibañez's original simulation for a slightly different purpose than it was set up for, as mentioned in §7.2. Sierra-Santibañez was researching how concept and words for connectives could emerge in some community of speakers. I am looking for how often the meanings of connectives are potentially useful for discrimination.

The discrimination games run in Sierra-Santibañez's original algorithm had the following properties:

- $N_D = 2$. There are 2 dimensions in the meaning space.
- $N_R = 4$. There are 4 regions in each meaning space dimension.
- $N_T \in \{1,2,3\}$. The number of topic objects has an equal chance of being either 1, 2 or 3.
- $N_B = 2$. There are 2 background objects.

Running 10,000 discrimination games using Sierra-Santibañez's original settings reveals the following frequency distribution:

TRUE	2060	\leftrightarrow	713	\vee	101	
\neg	3170	$\underline{\vee}$	1210	\leftarrow	35	
		AM	0	\rightarrow	44	
\wedge	60	MA	0	$ $	14	
BUT	128	NM	0			
ALTHOUGH	99	MN	0			Total 8697
\downarrow	1063					<i>Non-discriminable</i> 1303

From this distribution the following observations can be made:

- A. The two-placed connectives AM, MA, NM and MN are never used.
- B. \neg is used 1,5 times more often than TRUE, and \downarrow is used 17 times more often than \wedge .
- C. Although very frequent in natural language, \wedge is hardly used, as is \vee .
- D. Instead, \downarrow , $\underline{\vee}$ and \leftrightarrow are by far the most common two-placed connectives.

The connectives mentioned in observation (A) are Gazdar and Pullum’s non-compositional connectives. The reason they do not emerge is because in Sierra-Santibañez’s algorithm, a speaker agent searches for a characterisation using a one-placed connective first. If such a characterisation is found, the speaker does not bother with searching any further for a two-placed connective that may do the trick as well. Now, there is no case in which any of the non-compositional connectives could pick out some set that could not be picked out using an expression with just TRUE and \neg , because each of the non-compositional connectives is logically equivalent to such an expression. This corresponds to the situation in natural language (see appendix 2): there are no holistic connectives that are equivalent to AM, MA, NM or MN.³³

The connective use described in observations (B), (C) and (D) on the contrary does not correspond to the typical situation in natural language. In order to seek explanations for these phenomena, we need to tweak the original settings of Sierra-Santibañez and explore what their effects are on connective frequency.

7.3.2 *Manipulating Sierra-Santibañez’s original settings using a new algorithm*

We have been able to explain observation (A) easily from the properties of Sierra-Santibañez’s original algorithm. In order to find explanations for the other observations, we need to explore different combinations of the variables in the simulation: N_D , N_R , N_T and N_B . In order to manipulate these variables easily, I have rewritten the discrimination game algorithm in HyperTalk³⁴.

Other than allowing for easier manipulation of the simulation settings, we are now in a position to reconsider some of the built-in preferences of the original Prolog algorithm. Since we are using the algorithm to a slightly different purpose than it was originally meant for, some of these preferences may no longer be justifiable.

The discrimination simulations in the HyperTalk algorithm proceed as follows. Each simulation has a particular specification for N_D , N_R , N_B and N_T . First, *all* the possible objects using this particular number of dimensions and regions are generated. The number of possible objects equals $N_R^{N_D}$. The actual names of the dimensions and the regions into which they are divided are of no importance; the program just provides abstract symbol for them, e.g. dimensions X, Y and Z or regions a, b, c and d. Next, a random sample of 1,000 possible combinations of N_T topics and N_B background objects is generated using these objects. We have to limit ourselves to a sample since the number of possible combinations, given by $N_R^{N_D(N_T+N_B)}$, becomes too large quickly. The discrimination phase then proceeds by checking, for each pre-defined connective, for which one of the 1,000 possible combinations of topics and background objects the connective is true of the topic and false of the background.

The first change from Sierra-Santibañez’s Prolog code is that the new algorithm does an *exhaustive search*. The original algorithm stops after one discriminating connective has been found. That one connective is essentially a random pick of the possible connectives that could work in a particular discrimination game: a random set of arguments is chosen, after which a deterministic procedure leads to a connective. This makes sense in a simulation in which the goal of the agents is just to communicate. However if we want to know how frequently each connective could be used *potentially*, we need to continue searching until all the connectives that would discriminate in a particular round of the discrimination game have been found.

³³ Interestingly, Sierra-Santibañez reports in her (2001) paper that *all* 16 two-placed connectives plus TRUE and \neg arise during the simulation, even the non-compositional ones. This is because she simulated the emergence of one-placed and two-placed connectives separately. If a speaker needs a one-placed connective, but cannot use one, the non-compositional connectives fill the gap (Josefina Sierra-Santibañez p.c.).

³⁴ HyperTalk is a scripting language originally used in an Apple Macintosh authoring environment called HyperCard, which was very popular during the 1980’s and 1990’s. HyperCard has been discontinued by its developer Apple Computer Inc., but HyperTalk is still supported as part of a HyperCard clone called Revolution, developed by Runtime Revolution Ltd. of Edinburgh.

A second property of the Sierra-Santibañez simulation is a preference for one-placed over two-placed connectives. In this thesis I want to focus on whether two-placed connectives are useful in particular situations. *Prima facie* it may seem safe to say that a two-placed connective is not needed if an expression using TRUE or \neg would be logically equivalent. If we think more carefully, however, we need to wonder what the basis for that preference would be. Invoking least effort as an argument here would be inappropriate, since we are only dealing with the *semantic* meanings of connectives rather than their surface forms, and the principle of least effort has only been postulated for actual language. Instead we would just have to assume here that in mental logic, too, a preference somehow exists for one-placed over two-placed connectives.³⁵

Since we want to build as few assumptions as possible into our simulation, I will initially research the behaviour of the two-placed connectives without a preference for one-placed connectives. Possibly the simulation would reveal a frequency distribution leading to connective morphology with a realistic share of one-placed and two-placed connectives, even without such a preference.

Another preference present in Santibañez's algorithm is that for the connective with the least possible number of DNF-disjuncts (cf. §4.6.2). Again, this preference is difficult to justify from empirical data of the human mind. There is some evidence for a correlation between processing difficulty and number of truth conditions but this evidence, discussed in §6.4, is sketchy at best. Therefore my algorithm shall not necessarily prefer connectives with less DNF-disjuncts. If a set of objects can be discriminated using both $\underline{\vee}$ and \vee , for instance, both connectives will be equally acceptable, even though $\underline{\vee}$ could be said to be more specific than \vee .

Another choice to be made is whether I allow discrimination games to use the same object twice. By the 'same' object is meant an object with exactly the same properties.³⁶ Having the same object both in the topic and the background makes all discrimination impossible. This is not equally true for the cases where the same object is represented twice in either topic or background. Disallowing the same object to be used twice would make some combinations of settings impossible; for instance $N_R = 2$, $N_D = 2$ and $N_T = 8$. Since with $N_R = 2$ and $N_D = 2$ only 4 different objects are possible, it would be impossible to have eight topic objects. Since this happens to be one of the settings I have been meaning to use in the simulation runs, I have decided to allow using the same object multiple times in the discrimination games.

7.3.3 *Methods of statistical analysis of the data*

After running the simulations in HyperTalk, I have used the SPSS statistics package³⁷ to analyse the effects of the settings N_D , N_R , N_T , N_B on the frequencies of the connectives. SPSS is also useful for uncovering interactions between the effects of the settings. The results from the first run prompted for a second run with some different settings. Both runs and their details are covered in §7.3.4 and §7.3.5.

I have treated each of the simulations as a separate case, specified for N_D , N_R , N_T , N_B and for the number of times each connective was used. N_D , N_R , N_T , N_B are the independent variables upon which the frequencies of the connectives depend. I have analysed of the influence of these independent variables on the connective frequencies by plotting the relevant data.

Because the results of the discrimination game simulations are not real statistical data, I have refrained from using actual tests of statistical significance. Of each possible combination of the independent variables, only one case is present in my data. This would rule out any statistically significant results from this data if it had been experimentally obtained. However, several runs of the

³⁵ Such *mental* least-effort assumptions have been made in the literature. For instance Vogt (2004) has fitted robot agents with a tendency to minimise cognitive cost by preferring general concepts over specific ones. Interestingly, this assumption gave rise to the emergence of Zipf's law of word length and frequency in the agents' vocabularies.

³⁶ That two objects sharing the same properties are the 'same' objects by definition is a philosophical position famously taken by Leibniz (see Forrest 2002). This philosophical debate lies out of the scope of this dissertation, however.

³⁷ As its name implies, SPSS or the 'Statistics Package for the Social Sciences' is the standard computer programme used for data analysis in the social sciences. It is developed by SPSS Inc. of Chicago, IL. The MacOS version of SPSS is a horrible port that crashes and corrupts data continually, increasing my frustration levels to great heights.

simulation have shown that its outcome is almost completely deterministic: each run yields the same results, give or take 1 case. The only factor introducing variance is the random selection of topics and background objects in the 1,000 discrimination games.

SPSS, a *Statistical Package for the Social Sciences*, assumes the simulation data are in fact empirical, stochastic data from such social science experiments. However, the outcome of one run of a deterministic procedure is as significant as an infinite number of runs of a field experiment in the social sciences. The results on statistical significance that SPSS returns will be meaningless; instead, I will consider all the main and interactions effect found for N_D , N_R , N_T , N_B to be significant and not due to chance.

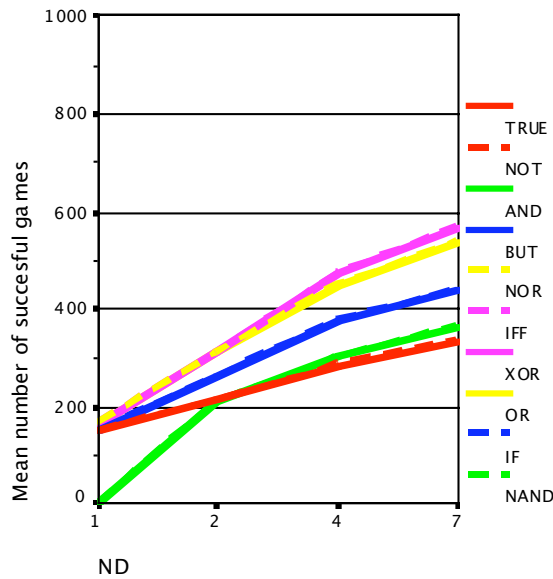
7.3.4 Results from run 1: more meaning space dimensions means easier discrimination

In run 1, I have done 256 simulations of 1,000 discrimination games each using the HyperTalk algorithm. The 256 games represent all the possible combinations of the following settings: $N_D \in \{1,2,4,7\}$; $N_R \in \{1,2,4,7\}$; $N_T \in \{1,2,4,8\}$; $N_B \in \{1,2,4,8\}$. This preliminary run revealed a number of properties of the diverse connectives that prompted me to do a second run with changes.

Firstly, I found that some connectives consistently had identical frequencies and 100 % overlap with other connectives. These were AM, MA, NM, MN, ALTHOUGH and \leftarrow . The first four are equivalent to TRUE and \neg , as noted before in observation (A). As for ALTHOUGH and \leftarrow , the data from run 1 has shown that these connectives always fully overlap with BUT and \rightarrow , respectively, and vice versa. This is a consequence of the fact that these non-commutative connectives mirror each other's argument positions: $P \text{ BUT } Q \equiv Q \text{ ALTHOUGH } P$ and $P \rightarrow Q \equiv Q \leftarrow P$. So, in every case in which ALTHOUGH or \leftarrow can be used, BUT or \rightarrow can be used, too, if the arguments are reversed.³⁸

Secondly, for none of the connectives does an interaction effect exist between N_D and either N_R , N_T or N_B . The main effect for N_D is depicted in fig. 2 below.³⁹

Figure 2. Main effect of number of meaning space dimensions (N_D) on the number of times $N_{\text{disc}}[C]$ in which each connective C successfully discriminates topic from background.



³⁸ Strangely, there is a small difference in number of occurrence for BUT and ALTHOUGH in Sierra-Santibañez's original algorithm, and similarly for \rightarrow and \leftarrow . I had no explanation for this result at the time of writing.

³⁹ Every time a graph is presented with a main effect for some independent variable X , this means that for each value of X , the mean number of successful games over all the different settings for the other independent variables is presented.

There is positive main effect of N_D for all the connectives: with an increase in number of dimensions, the number of cases in which any connective can successfully discriminate increases. This intuitively makes sense: each of the meaning space dimensions acts as a separate ‘opportunity’ for applying a particular connective. The more meaning space dimensions, the more elaborately specified the objects can be, and the larger the chances are that an object can be accurately described using some connective.

The effect is slightly stronger for the pairs \leftrightarrow / \vee and \vee / \downarrow while less strong for \wedge / \mid . However, as N_D increases the lines start running more parallel, i.e. $\Delta N_{\text{disc}} / \Delta N_D$ for each connective converges. Most importantly, nowhere does a change in N_D change the relative order of successfulness of the connectives. Therefore there is no way in which the variable N_D could make a difference as to which connective works best in a given situation. Since there is also no interaction effect between N_D and either N_R , N_T or N_B , for any of the individual connectives, we can ignore N_D in the rest of our analysis and can suffice with keeping the setting for N_D equal across all simulation runs.⁴⁰

Moving on to the effect of N_R , I found that the number of successful games using any connective is 0 if $N_R = 1$. This must be the case since if a particular property of an object (say COLOUR) can have only one value (say *red*), that particular property has no discriminatory value: all objects are red by definition. We can eliminate the setting $N_R = 1$ from the data.

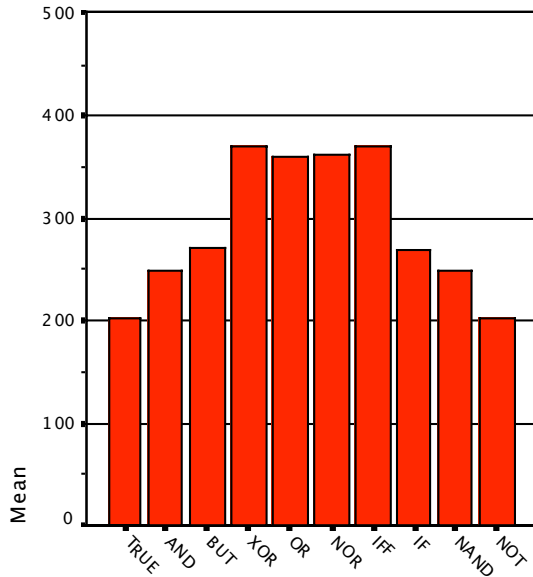
7.3.5 Results from run 2: the effects of number of regions, topics and background objects interact

In run 2 there were 100 simulation of 1,000 discrimination games, where $N_D = 3$, $N_R \in \{2,3,5,7\}$, $N_T \in \{1,2,3,5,7\}$ and $N_B \in \{1,2,3,5,7\}$. These settings reflect several adjustments made since run 1. $N_D = 3$ across all runs, since the results from run 1 showed that dimensionality does not interact with the effect of the other variables. $N_R = 1$ is removed, since with this setting no successful discrimination can occur using any connective. Since keeping N_D constant reduces the number of—computation intensive—simulations, I was able to increase the number of data points by introducing some more possible settings for N_R , N_T and N_B .

⁴⁰ The connectives \wedge and \mid show some aberrant behaviour: $N_{\text{disc}}[\wedge]$ and $N_{\text{disc}}[\mid]$ are both 0 if $N_D = 0$. This must be the case because if $r_1 \neq r_2$, $\wedge(X_J(r_1), X_J(r_2))$ is a contradiction while $\mid(X_J(r_1), X_J(r_2))$ is a tautology. If $r_1 = r_2$, $\wedge(X_J(r_1), X_J(r_2))$ is equivalent to $\text{TRUE}(X_J(r_1))$ or $\text{TRUE}(X_J(r_2))$ while $\mid(X_J(r_1), X_J(r_2))$ is equivalent to $\neg(X_J(r_1))$ or $\neg(X_J(r_2))$. In these latter cases \wedge and \mid cannot discriminate because in run 1, one-placed TRUE and \neg were always preferred over two-placed \wedge and \mid (cf. §7.3.2).

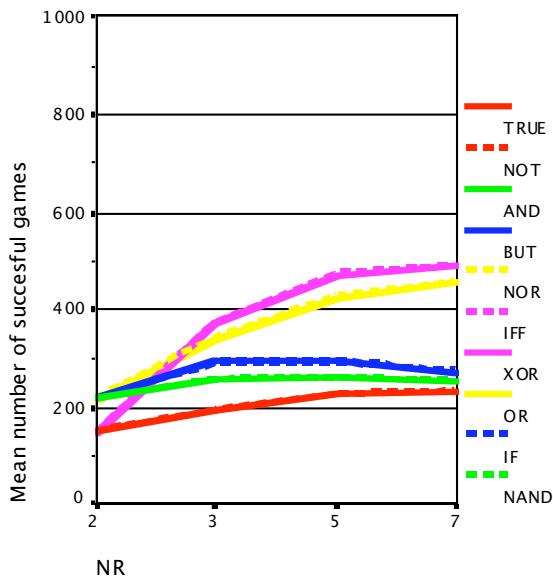
The graph in fig. 3 below depicts the mean number of discrimination games each connective was useful in across all settings.

Figure 3. Mean number of times $N_{disc}[C]$ in which each connective C successfully discriminates topic from background in run 2.



Surprisingly, TRUE and \neg , on the far left and right of the graph, are the connectives used least. Also, \wedge is infrequent. The most frequently applicable connectives are \vee and \Leftrightarrow . If will defer discussing the causes of this particular frequency distribution to §7.3.8. Using the results of run 2, I shall now investigate the main effect for number of regions per meaning space dimension, presented in fig. 4 below.

Figure 4. Main effect of number of regions per each meaning space dimension (N_R) on the number of times $N_{disc}[C]$ in which each connective C successfully discriminates topic from background.



The main effect for the number of regions is less clear than that for number of meaning space dimensions. Most noticeably, the 10 connectives fall into 5 pairs with respect to their average frequency, namely TRUE / \neg , \wedge / \mid , \vee / \downarrow , BUT / \rightarrow and $\underline{\vee}$ / \leftrightarrow . These connective pairs are each other's negations:

$$\begin{array}{lll}
 \text{TRUE}(P) \equiv \neg \neg(P) & \wedge & \neg(P) \equiv \neg \text{TRUE}(P) \\
 P \wedge Q \equiv \neg(P \mid Q) & \wedge & P \mid Q \equiv \neg(P \wedge Q) \\
 P \vee Q \equiv \neg(P \downarrow Q) & \wedge & P \downarrow Q \equiv \neg(P \vee Q) \\
 P \text{ BUT } Q \equiv \neg(P \rightarrow Q) & \wedge & P \rightarrow Q \equiv \neg(P \text{ BUT } Q) \\
 P \underline{\vee} Q \equiv \neg(P \leftrightarrow Q) & \wedge & P \leftrightarrow Q \equiv \neg(P \underline{\vee} Q)
 \end{array}$$

If $N_R = 2$, an even larger number of connectives have the same average number of successful games: these are the pairs TRUE / \neg and $\underline{\vee}$ / \leftrightarrow on one hand and \wedge / \mid , \vee / \downarrow , and BUT / \rightarrow on the other hand. The connective pairs \vee / \downarrow and $\underline{\vee}$ / \leftrightarrow increase in frequency fairly rapidly as N_R goes up, while other connective pairs' frequencies increase only slowly or even decrease slightly with an increase of N_R (TRUE / \neg , \wedge / \mid and BUT / \rightarrow).

A similar 'pairing up' effect is found in the main effects of number of topics N_T and number of background objects N_B . These are presented in fig. 5 below.

Figure 5a. Main effect of the number of topic objects N_T on the number of times $N_{\text{disc}}[C]$ in which each connective C successfully discriminates topic from background.

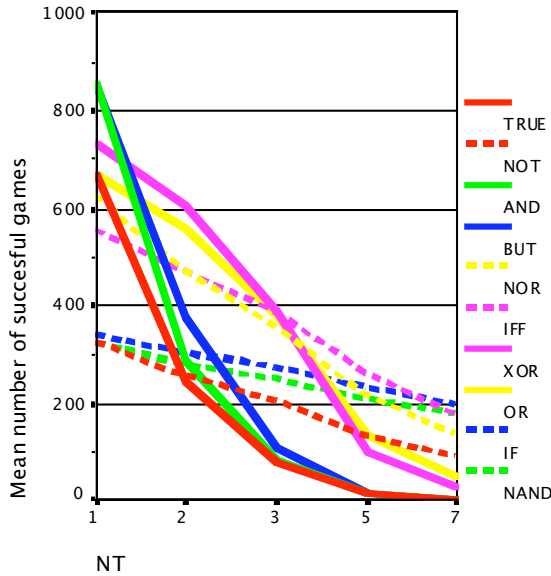
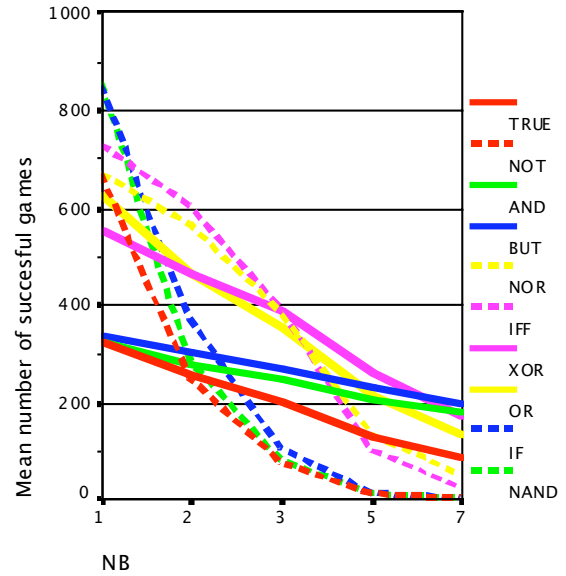


Figure 5b. Main effect of the number of background objects N_B on the number of times $N_{\text{disc}}[C]$ in which each connective C successfully discriminates topic from background.



As the number of topic objects increases, the number of cases in which each connective discriminates decreases. The larger the set of topics that have to be picked out, the more difficult it gets to find an expression that describes that set of topics, if the expression has to be limited to one connective with two properties as its arguments. Likewise, the more background objects there are, the more difficult it gets to find an expression that sets any topic apart from that background. All this is true no matter what connective is used in the expression.

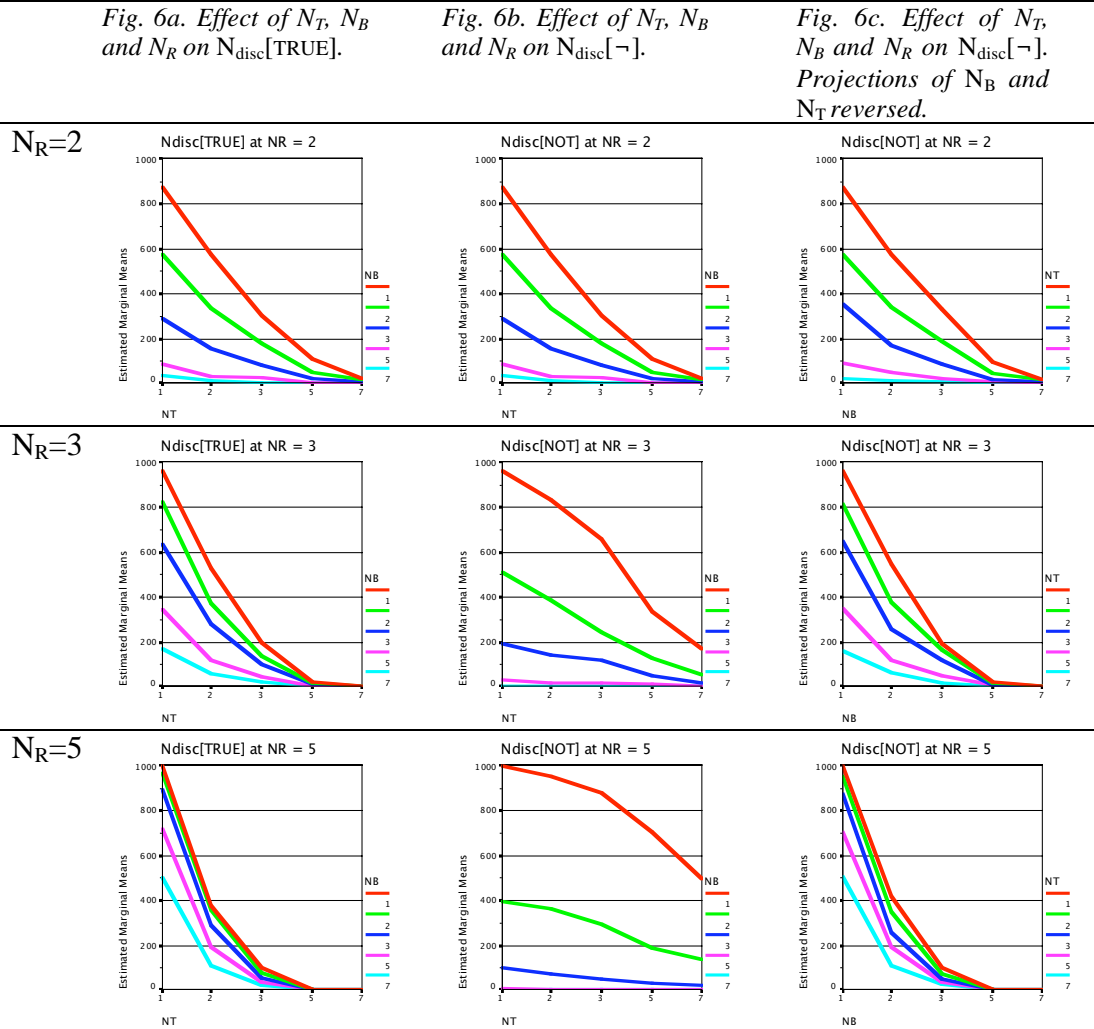
The decrease in number of successful games is not equally large for each connective, however. Some connectives lose their use much faster than others as N_T or N_B increases. For instance, if $N_T = 1$, \neg discriminates the topic from the background $N_{\text{disc}}[\neg] = 320$ times on average, while TRUE is used

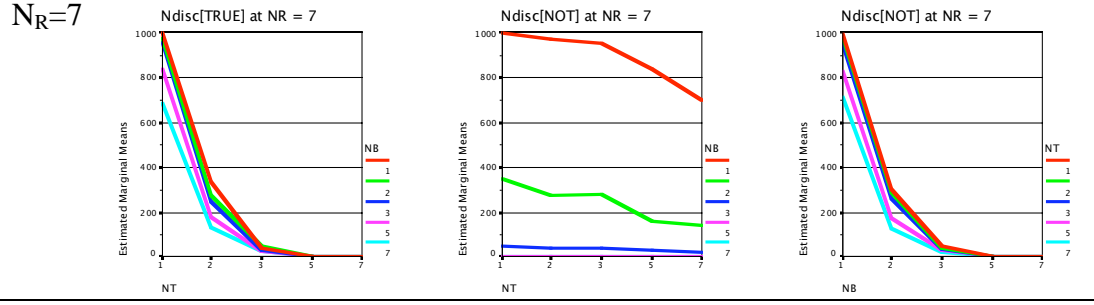
680 times. If N_T increases to 2, however, $N_{\text{disc}}[\text{TRUE}]$ sinks by 450 to about 230, while $N_{\text{disc}}[\neg]$ goes down just by 40, so \neg is used 240 times on average, more often than TRUE is used. So, although both TRUE and \neg discriminate less well as the number of topic objects increases, \neg is affected less by this increase, and so \neg becomes more successful than TRUE once $N_T \geq 2$.

Interestingly, the opposite situation is found in graph 5b depicting the change in N_{disc} as N_B increases rather than N_T . With $N_B = 1$, $N_{\text{disc}}[\neg]$ starts out high, and above $N_{\text{disc}}[\text{TRUE}]$, but as N_B gets larger than 2, $N_{\text{disc}}[\neg]$ plummets below the level of $N_{\text{disc}}[\text{TRUE}]$. Here, the two connectives TRUE and \neg mirror each other's decrease in successfulness as a consequence of an increase of N_T and N_B , respectively. The same pattern is found with the connective pairs \wedge / \mid , \vee / \downarrow , BUT / \rightarrow and $\underline{\vee} / \leftrightarrow$ as well. These pairs of connectives are, like TRUE and \neg , each other's negations. They are the same pairs of connectives that were paired up in the graph depicting the main effect of N_R (fig. 4). In the graphs in fig. 5, the second member of these pairs is marked with a dotted line in the same colour in which the first member is marked.

In fig. 6 below we find the interaction effects between N_T , N_B and N_R for one of these five pairs, TRUE / \neg .

Figure 6. Interaction effects of number of topic objects N_T , number of background objects N_B and number of regions N_R on the number of times TRUE and \neg are successfully used ($N_{\text{disc}}[\text{TRUE}]$ and $N_{\text{disc}}[\neg]$).





The graphs for \neg and TRUE in fig. 6a and fig. 6b are identical when $N_R = 2$. However, as N_R increases, the number of successful games for TRUE and \neg starts to diverge. When N_R reaches high values, TRUE yields a large number of successful games when $N_T < 2$, but as N_T increases, $N_{\text{disc}}[\text{TRUE}]$ quickly approaches 0. The situation for \neg is the other way around: $N_{\text{disc}}[\neg]$ is high when $N_B = 1$, medium high when $N_B=2$ and approaches 0 when $N_B > 2$. $N_{\text{disc}}[\neg]$ is hardly influenced by N_T , while $N_{\text{disc}}[\text{TRUE}]$ does responds very little to changes in N_B .

TRUE and \neg seem to mirror each other's behaviour: TRUE reacts to changes in N_T the same way as \neg reacts to changes in N_B , and vice versa. Indeed, they mirror each other perfectly, which can be seen if we flip the way N_B and N_T are projected in the graph for $N_{\text{disc}}[\neg]$. If N_B is projected along the x-axis while N_T is represented with separate lines, the graph for $N_{\text{disc}}[\neg]$ looks exactly like the one for $N_{\text{disc}}[\text{TRUE}]$, as can be seen by comparing figures 6a and 6c. This also means that $N_{\text{disc}}[\text{TRUE}] = N_{\text{disc}}[\neg]$ if $N_T = N_B$, which can also be verified from the graphs in fig. 4.

Comparable graphs for all the 10 connectives investigated are presented in appendix 4. From these graphs, we notice that the connectives fall into two groups with behaviours that are similar to either TRUE (the *positive* connectives) or \neg (the *negative* connectives). The positive connectives do well with a low number of topics, but quickly approach zero successful games as N_T gets higher. An increase in the number of background objects does not affect the success of these connectives much.

The negative connectives are the negations of the positive connectives. Their behaviour is characterised by a high number of successful games when N_B is low, which quickly approaches zero as N_B increases. An increase in the number of topic objects does not affect the success of these connectives much. For all of the connectives that are each other's negations, the graphs are identical if the projection axes of N_B and N_T are flipped for one of them.

Within the two groups, connectives can be more or less extreme in their behaviour. Along this dimension of extremity of behaviour, both the positive and negative groups of connectives can be dichotomised into moderate and extreme members of that group. Extremely positive connectives are \wedge , BUT and TRUE: their usefulness collapses very quickly once N_T gets larger than 1. Moderately positive connectives are \vee and \vee : like the extremely connectives, the number of times they discriminate diminishes once N_T gets larger than 1, only not as fast. Extremely negative connectives are \mid , \rightarrow and \neg , while moderately negative connectives are \leftrightarrow and \downarrow . These connectives start yielding fewer successful games once N_B increases beyond 1: the extreme ones very rapidly, the moderate ones less so.

I have presented the positive and negative categories of connectives in table 2 below, with descriptions of their behaviour and example graphs that are representative of the graphs for the other connectives in that category. As mentioned before, graphs for all the connectives can be found in appendix 4.

Table 2. Groups of connectives based on their behaviours under various settings of N_T and N_B , assuming $N_R > 2$.

Category	Positive	Negative
Behaviour	High number of successful games when N_T is low, quickly approaching 0 as N_T increases; N_B has little influence.	High number of successful games when N_B is low, quickly approaching 0 as N_B increases; N_T has little influence.
Extreme example graphs: \wedge (positive) and \neg (negative), $N_R = 7$		
Moderate example graphs: \vee (positive) and \downarrow (negative), $N_R = 7$		
Extreme ↓ Moderate	\wedge —————→ BUT/ALTHOUGH —————→ TRUE —————→ \vee —————→ \vee —————→ negation	\neg → / ← ⌋ ↔ ↓

The dichotomy between negative and positive gets more pronounced as N_R increases, but it is mostly absent when $N_R = 2$. In this case, the connectives show fairly similar behaviour that straddles a middle ground between the extreme behaviours described for the positive and negative groups. The connective groups TRUE / \neg , as well \wedge / BUT / \downarrow , \neg / \rightarrow / \vee and \leftrightarrow / $\underline{\vee}$ even yield identical numbers of successful games if $N_R = 2$.

The analysis of the interaction effects of N_R , N_T and N_B has left us with several explananda:

- I. Why do the connectives in the *positive* group yield a high number of successful games when N_T is low, quickly approaching 0 as N_T increases, while N_B has little influence?
- II. Why do the connectives in the *negative* group yield a high number of successful games when N_B is low, quickly approaching 0 as N_B increases, while N_T has little influence?
- III. Why do the connectives in both groups mirror each other's response to N_T and N_B ?
- IV. Why is the dichotomy in a positive and a negative group small to absent when $N_R = 2$, and does it get more pronounced as N_R increases?

In next paragraph I shall provide an explanation for these results.

7.3.6 Explaining the 'positive' and 'negative' reactions of connectives to N_T and N_B

We have noticed that the connectives can be divided into 5 pairs. For each of the pairs, one member works relatively well when there are many topics, while the other works better with many background objects. Let us investigate one of these pairs, the connectives TRUE / \neg . The key to explaining the two groups is the observation that the dichotomy is small to absent when $N_R = 2$, while it gets more pronounced as N_R increases. If a meaning space is divided into more than 2 regions, there are more

worlds in which $\neg(X)$ is true than worlds in which $\text{TRUE}(X)$ is true.

We can quantify this difference in number of worlds. Assume one meaning space dimension X divided into 4 regions $R_X = \{a, b, c, d\}$ and one object J . This means there are four possible worlds, namely those in which respectively $X_J = a$, $X_J = b$, $X_J = c$ and $X_J = d$.⁴¹ For any object J and for any region $r \in R_X$, $\text{TRUE}(X_J(r))$ is true if and only if $X_J = r$, whereas $\neg(X_J(r))$ is true if and only if $X_J \in R - r$, i.e. the complement of r with respect to R_Q , the equivalent of negation in set theory. For instance, $\text{TRUE}(X_J(a))$ is true if and only if $X_J = a$, while $\neg(X_J(a))$ is true if and only if $X_J = b$, $X_J = c$ or $X_J = d$. Thus, given one meaning space dimension X , one object J with property $X_J(x)$ can be described using $\text{TRUE}(X_J(a))$ in $N_{\text{true}}[\text{TRUE}(X_J(a))] = 1$ world and using $\neg(X_J(a))$ in a number of worlds equal to the cardinality of the complement of r with respect to R_Q , i.e. in $N_{\text{true}}[\neg(X_J(a))] = N_R - 1$ cases, where N_R is the number of regions in X .

Two objects J_1 and J_2 can only be described with $\text{TRUE}(X_J(a))$ if and only if $X_{J1} = a \wedge X_{J2} = a$, but using $\neg(X_J(a))$ if and only if $X_{J1} \in \{b, c, d\} \wedge X_{J2} \in \{b, c, d\}$, i.e. in $3^2 = 9$ worlds. If there are **three** objects, $N_{\text{true}}[\neg(X_J(a))]$ adds up to $3^3 = 27$, while $\text{TRUE}(X_J(a))$ remains true in only one world. Generally, for any number of objects N_J , $\text{TRUE}(X_J(r))$ can be true in $N_{\text{true}}[\text{TRUE}(X_J(r))] = 1$ world, while $\neg(X_J(r))$ can be true in $N_{\text{true}}[\neg(X_J(r))] = N_R - 1^{N_J}$ worlds.

The asymmetry we found for \neg and TRUE is magnified in the case of \downarrow and \wedge . Some complications arise if we want to investigate $N_{\text{true}}[\wedge]$ and $N_{\text{true}}[\downarrow]$, because if there is only one meaning space dimension, not all combinations of regions are possible: for instance, $\wedge(X_J(r_1), X_J(r_2))$ is a contradiction if $r_1 \neq r_2$, while it is equivalent to $\text{TRUE}(X_J(r_1))$ or $\text{TRUE}(X_J(r_2))$ if $r_1 = r_2$. However, in the case of two dimensions X and Y , $\wedge(X_J(r_1), Y_J(r_2))$ is not a contradiction if $r_1 \neq r_2$. Using different dimensions as arguments leads to different behaviours of the connectives (see also note 40).

We will circumvent these problems for now and assume two meaning space dimensions X and Y , both divided into 4 regions $R_X = R_Y = \{a, b, c, d\}$. In this case, there is one way in which $\wedge(X_J(a), Y_J(a))$ can be true, namely if and only if $X_J = a \wedge Y_J = a$, but $\downarrow(X_J(a), Y_J(a))$ is true if and only if $X_J \in \{b, c, d\} \wedge Y_J \in \{b, c, d\}$, which is true in $3^2 = 9$ worlds. If we consider two objects J_1 and J_2 , $\wedge(X_J(a), Y_J(a))$ is true if and only if $X_{J1} = a \wedge X_{J2} = a \wedge Y_{J1} = b \wedge Y_{J2} = b$, while $\downarrow(X_J(a), Y_J(a))$ is true if and only if $X_{J1} \in \{b, c, d\} \wedge Y_{J1} \in \{b, c, d\} \wedge X_{J2} \in \{b, c, d\} \wedge Y_{J2} \in \{b, c, d\}$, which is true in $(N_R - 1)^{2N_J} = 3^4 = 81$ worlds. The formula for $N_{\text{true}}[\downarrow(X_J(r_1), Y_J(r_2))]$, $(N_R - 1)^{2N_J}$, can be rewritten as $(N_R^2 - 2N_R + 1)^{N_J}$, which is the formula I shall use from now on for more uniformity across the formulae for the different connectives.

The connectives **BUT** and **ALTHOUGH** are in between \downarrow and \wedge in number of worlds in which they are true. **BUT**($X_J(a), Y_J(a)$) is true if and only if $X_J = a \wedge Y_J \in \{b, c, d\}$, which is true in $N_R - 1 = 3$ worlds. Likewise **ALTHOUGH**($X_J(a), Y_J(a)$) is true in $N_R - 1 = 3$ worlds as well. If $N_J = 2$, **BUT**($X_J(a), Y_J(a)$) is true if and only if $X_{J1} = a \wedge Y_{J1} \in \{b, c, d\} \wedge X_{J2} = a \wedge Y_{J2} \in \{b, c, d\}$, which is true in $(N_R - 1)^{N_J} = 3^2 = 9$ worlds.

Finally, the number of worlds in which TRUE and \neg discriminate changes as well if two dimensions rather than one are introduced. $\text{TRUE}(X_J(a))$ is still true if and only if $X_J = a$, but at the same time it can be the case that $Y_J = a$, $Y_J = b$, $Y_J = c$ or $Y_J = d$, so here $\text{TRUE}(X_J(a))$ is true in $1 * N_R = 4$ worlds. In the case of two topics, $\text{TRUE}(X_{J1}(a))$ true if and only if $X_{J1} = a \wedge X_{J2} = a \wedge Y_{J1} \in \{a, b, c, d\} \wedge X_{J2} \in \{a, b, c, d\}$. So, if $N_D = 2$, $\text{TRUE}(X_J(a))$ is true in $1 * N_R^{N_J} = 4^2 = 16$ worlds. Likewise, we multiply the formula we obtained before for $N_{\text{true}}[\neg(X_J(r))]$ for $N_D = 1$ by $N_R^{N_J}$ if we want a formula valid for $N_D = 2$. So if $N_D = 1$, $N_{\text{true}}[\neg(X_J(r))] = (N_R - 1)^{N_J}$, and for $N_D = 2$, $N_{\text{true}}[\neg(X_J(r))] = N_R^{N_J} (N_R - 1)^{N_J} = (N_R^2 - N_R)^{N_J}$.

This brings us the following formulas for the number of worlds in which TRUE , \neg , \wedge , **BUT**, \downarrow are true, assuming $N_D = 2$.

⁴¹ X_J represents the property X for object J .

$$\begin{aligned}
N_{\text{true}}[\text{TRUE}(X_J(r))] &= N_R^{\text{NJ}} \\
N_{\text{true}}[\neg(X_J(r))] &= (N_R^2 - N_R)^{\text{NJ}} \\
\\
N_{\text{true}}[\wedge(X_J(r_1), Y_J(r_2))] &= 1 \\
N_{\text{true}}[\text{BUT}(X_J(r_1), Y_J(r_2))] &= (N_R - 1)^{\text{NJ}} \quad (= N_{\text{true}}[\text{ALTHOUGH}(X_J(r_1), Y_J(r_2))]) \\
N_{\text{true}}[\downarrow(X_J(r_1), Y_J(r_2))] &= (N_R^2 - 2N_R + 1)^{\text{NJ}}
\end{aligned}$$

If more than two dimensions are introduced, these numbers need to be multiplied by a factor involving N_D and N_R to account for all the other values the dimensions could take on that are not used in the representation. I shall not derive those formulae here.

For all the other two-placed connectives C , we find $N_{\text{true}}[C(X_J(r_1), Y_J(r_2))]$ by summing the values of $N_{\text{true}}[\wedge(X_J(r_1), Y_J(r_2))]$, $N_{\text{true}}[\text{BUT}(X_J(r_1), Y_J(r_2))]$ or $N_{\text{true}}[\downarrow(X_J(r_1), Y_J(r_2))]$, depending on which of these connectives are part of the DNF-representation of the desired connective. For instance, since $P \leftrightarrow Q$ is true if and only if $(P \wedge Q) \vee (P \downarrow Q) \equiv 1$, the number of worlds in which $P \leftrightarrow Q$ is true equals the sum of the number of worlds in which $P \wedge Q$ and $P \downarrow Q$ are true. The summation is done before we take the number of objects into account by elevating the formulae to the power of N_J .⁴²

Below, I derive N_{true} for all two-placed connectives. For purposes of notational legibility, $N_{\text{true}}[C(X_J(r_1), Y_J(r_2))]$ is from now on shortened to $N_{\text{true}}[C(J)]$, meaning the number of worlds in which an expression with connective C is true of a set of objects J , assuming there are two meaning space dimensions X and Y , X is used for the first argument of the expression, Y for the second argument, and two different meaning space regions are used for each dimension.⁴³

$$\begin{aligned}
N_{\text{true}}[\wedge(J)] &= 1 \\
N_{\text{true}}[\text{BUT}(J)] &= N_{\text{true}}[\text{ALTHOUGH}(J)] &= (N_R - 1)^{\text{NJ}} \\
N_{\text{true}}[\downarrow(J)] &= (N_R^2 - 2N_R + 1)^{\text{NJ}} \\
\\
N_{\text{true}}[\leftrightarrow(J)] &= N_{\text{true}}[\wedge(J)] + N_{\text{true}}[\downarrow(J)] &= (1 + N_R^2 - 2N_R + 1)^{\text{NJ}} &= (N_R^2 - 2N_R + 2)^{\text{NJ}} \\
N_{\text{true}}[\underline{\vee}(J)] &= 2N_{\text{true}}[\text{BUT}(J)] &= (2(N_R - 1))^{\text{NJ}} &= (2N_R - 2)^{\text{NJ}} \\
\\
N_{\text{true}}[\vee(J)] &= N_{\text{true}}[\wedge(J)] + 2N_{\text{true}}[\text{BUT}(J)] &= (1 + 2(N_R - 1))^{\text{NJ}} &= (2N_R - 1)^{\text{NJ}} \\
N_{\text{true}}[\rightarrow(J)] &= N_{\text{true}}[\wedge(J)] + N_{\text{true}}[\text{BUT}(J)] + N_{\text{true}}[\downarrow(J)] &= (1 + N_R - 1 + N_R^2 - 2N_R + 1)^{\text{NJ}} &= (N_R^2 - N_R + 1)^{\text{NJ}} \\
N_{\text{true}}[\mid] &= 2N_{\text{true}}[\text{BUT}(J)] + N_{\text{true}}[\downarrow(J)] &= (2(N_R - 1) + N_R^2 - 2N_R + 1)^{\text{NJ}} &= (N_R^2 - 1)^{\text{NJ}}
\end{aligned}$$

The connectives in the negative group (\mid , \rightarrow , \neg , \leftrightarrow and \downarrow) all contain the term N_R^2 in the formula deriving the number of worlds in which they are true. This terms stems from the inclusion of \downarrow in the

⁴² That this approach is correct is confirmed when we try to derive $N_{\text{true}}[\text{AM}]$, $N_{\text{true}}[\text{MA}]$, $N_{\text{true}}[\text{NM}]$ and $N_{\text{true}}[\text{MN}]$ in this way, i.e. the number of worlds in which AM, MA, NM and MN are true:

$$\begin{aligned}
N_{\text{true}}[\text{AM}] &= N_{\text{true}}[\text{MA}] = N_{\text{true}}[\wedge] + N_{\text{true}}[\text{BUT}] = (1 + N_R - 1)^{\text{NJ}} = N_R^{\text{NJ}} = N_{\text{true}}[\text{TRUE}] \\
N_{\text{true}}[\text{NM}] &= N_{\text{true}}[\text{MN}] = N_{\text{true}}[\downarrow] + N_{\text{true}}[\text{BUT}] = (N_R - 1 + N_R^2 - 2N_R + 1)^{\text{NJ}} = (N_R^2 - N_R)^{\text{NJ}} = N_{\text{true}}[\neg]
\end{aligned}$$

These four two-placed connectives, which are logically equivalent to TRUE and \neg , also turn out to have the same formulae for deriving N_{true} as those two connectives.

⁴³ Of course more elements of R_X and R_Y can be used to build an expression with any one or two-placed connective, and different combinations of dimensions can be used. The number of expressions possible using a two-placed connective equals $N_R^2 N_D^2$. In many cases the same connective can discriminate the same objects in several ways by using different arguments. For our research however, we are not really interested in how many ways one particular connective could discriminate one set of objects, but rather how many possible worlds (i.e. sets of topics and background objects) a particular connective can discriminate at least once, using any set of arguments. It gets very complicated to derive the relevant mathematical formulae that take all possible combinations of arguments into account. I shall leave these complications aside, since the formulae we derived thus far already explain the data well.

Disjunctive Normal Form of these connectives: each of them is true if both of its arguments are false. These are the non-confessional connectives of Gazdar and Pullum (see §4.2). The term N_R^2 makes N_{true} grow much quicker for the negative connectives than for the positive ones as N_J gets larger. So, negative connectives are true relatively more often of a large number of objects than of a small number of objects. Positive connectives on the other hand are true relatively more often of a small number of objects than of a large number of objects (explanandum I and II from §7.3.5).

From the number $N_{\text{true}}[C(X_J(r_1), Y_J(r_2))]$ of times that a connective C can be true of a set of objects, we can derive the number of times $N_{\text{disc}}[C(X_J(r_1), Y_J(r_2))]$ that C can discriminate a number of topics N_T from a number of background objects N_B .

In order to discriminate a topic from a background, a logical representation must be true of the topic objects and false of the background objects. For any connective C , $C(X_J(a))$ is false of some object if and only if $\neg C(X_J(a))$ is true. So, the number of cases $N_{\text{false}}[C(J)]$ in which $C(X_B(a))$ is false for some background object B equals the number of cases $N_{\text{true}}[\neg C(J)]$ in which the negation of C , $\neg C(X_B(a))$, is true for B .

For each of the cases in which some expression is true of the topic, there also are a number of cases in which the expression is false of the background. So generally, for any connective C , the number of cases in which C discriminates the topic from the background equals the number of worlds in which C is true of the topic multiplied by the number of worlds in which the negation of C is true of the background:

$$N_{\text{disc}}[C] = N_{\text{true}}[C(X_T(a))] * N_{\text{false}}[C(X_B(a))] = N_{\text{true}}[C(X_T(a))] * N_{\text{true}}[\neg C(X_B(a))]$$

If we apply this to all the 10 two-placed connectives we are researching, we get the following formulae:

Positive connectives

$$\begin{aligned} N_{\text{disc}}[\wedge] &= N_{\text{true}}[\wedge(T)] * N_{\text{true}}[\downarrow(B)] &= (N_R^2 - 1)^{NB} \\ N_{\text{disc}}[\text{BUT}] &= N_{\text{true}}[\text{BUT}(T)] * N_{\text{true}}[\rightarrow(B)] &= (N_R - 1)^{NT} * (N_R^2 - N_R + 1)^{NB} \\ N_{\text{disc}}[\text{TRUE}] &= N_{\text{true}}[\text{TRUE}(T)] * N_{\text{true}}[\neg(B)] &= N_R^{NT} * (N_R^2 - N_R)^{NB} \\ N_{\text{disc}}[\underline{\vee}] &= N_{\text{true}}[\underline{\vee}(T)] * N_{\text{true}}[\leftrightarrow(B)] &= (2N_R - 2)^{NT} * (N_R^2 - 2N_R + 2)^{NB} \\ N_{\text{disc}}[\vee] &= N_{\text{true}}[\vee(T)] * N_{\text{true}}[\downarrow(B)] &= (2N_R - 1)^{NT} * (N_R^2 - 2N_R + 1)^{NB} \end{aligned}$$

Negative connectives

$$\begin{aligned} N_{\text{disc}}[\downarrow] &= N_{\text{true}}[\downarrow(T)] * N_{\text{true}}[\vee(B)] &= (N_R^2 - 2N_R + 1)^{NT} * (2N_R - 1)^{NB} \\ N_{\text{disc}}[\leftrightarrow] &= N_{\text{true}}[\leftrightarrow(T)] * N_{\text{true}}[\underline{\vee}(B)] &= (N_R^2 - 2N_R + 2)^{NT} * (2N_R - 2)^{NB} \\ N_{\text{disc}}[\neg] &= N_{\text{true}}[\neg(T)] * N_{\text{true}}[\text{TRUE}(B)] &= (N_R^2 - N_R)^{NT} * N_R^{NB} \\ N_{\text{disc}}[\rightarrow] &= N_{\text{true}}[\rightarrow(T)] * N_{\text{true}}[\text{BUT}(B)] &= (N_R^2 - N_R + 1)^{NT} * (N_R - 1)^{NB} \\ N_{\text{disc}}[|] &= N_{\text{true}}[|(T)] * N_{\text{true}}[\wedge(B)] &= (N_R^2 - 1)^{NT} \end{aligned}$$

From the formulae thus derived we can explain the generalisations we made in our data analysis in §7.3.5. We notice that the formula for $N_{\text{disc}}[C]$ for some connective C always equals $N_{\text{disc}}[C]$ for the negation of C , except that N_T and N_B are reversed. This explains why the behaviours of the positive and negative connectives are **mirrored** with respect to their response to changes in N_T and N_B (explanandum III in §7.3.5).

As can be seen from the formulae, connectives in the **negative** group ($|$, \rightarrow , \neg , \leftrightarrow and \downarrow) increase rapidly in their number of successful games as N_T grows, and much more slowly with the increase of N_B . On the other hand, the connectives in the **positive** group (\wedge , BUT , TRUE , $\underline{\vee}$ and \vee) increase rapidly in their number of successful games as N_B grows, and much more slowly with the increase of N_T . Since the total number of possible cases also grows exponentially with N_T and N_B , the *proportion* of

cases in which positive connectives discriminate approaches 0 as N_T increases, and the *proportion* of cases in which negative connectives discriminate approaches 0 as N_B increases.

As for the fact that some connectives show similar behaviour if $N_R = 2$, if we supply $N_R = 2$ in the formulae above, we get:

$$\begin{array}{lll}
N_{\text{disc}}[\wedge] & = N_{\text{true}}[\wedge(T)] * N_{\text{true}}[\mid(B)] & = 3^{NB} \\
N_{\text{disc}}[\text{BUT}] & = N_{\text{true}}[\text{BUT}(T)] * N_{\text{true}}[\rightarrow(B)] & = 3^{NB} \\
N_{\text{disc}}[\text{TRUE}] & = N_{\text{true}}[\text{TRUE}(T)] * N_{\text{true}}[\neg(B)] & = 2^{NT} * 2^{NB} \\
N_{\text{disc}}[\underline{v}] & = N_{\text{true}}[\underline{v}(T)] * N_{\text{true}}[\leftrightarrow(B)] & = 2^{NT} * 2^{NB} \\
N_{\text{disc}}[v] & = N_{\text{true}}[v(T)] * N_{\text{true}}[\downarrow(B)] & = 3^{NT} \\
\\
N_{\text{disc}}[\downarrow] & = N_{\text{true}}[\downarrow(T)] * N_{\text{true}}[v(B)] & = 3^{NB} \\
N_{\text{disc}}[\leftrightarrow] & = N_{\text{true}}[\leftrightarrow(T)] * N_{\text{true}}[\underline{v}(B)] & = 2^{NT} * 2^{NB} \\
N_{\text{disc}}[\neg] & = N_{\text{true}}[\neg(T)] * N_{\text{true}}[\text{TRUE}(B)] & = 2^{NT} * 2^{NB} \\
N_{\text{disc}}[\rightarrow] & = N_{\text{true}}[\rightarrow(T)] * N_{\text{true}}[\text{BUT}(B)] & = 3^{NT} \\
N_{\text{disc}}[\mid] & = N_{\text{true}}[\mid(T)] * N_{\text{true}}[\wedge(B)] & = 3^{NT}
\end{array}$$

So we can derive that if $N_R = 2$, the number of successful discrimination games is determined solely by the number of cases for which a connective yields true.

$$\begin{array}{lll}
N_{\text{disc}}[\wedge] = N_{\text{disc}}[\text{BUT}] & = N_{\text{disc}}[\downarrow] & 1 \text{ disjunct in the DNF-representation} \\
N_{\text{disc}}[\text{TRUE}] = N_{\text{disc}}[\underline{v}] & = N_{\text{disc}}[\neg] = N_{\text{disc}}[\leftrightarrow] & 2 \text{ disjuncts in the DNF-representation} \\
N_{\text{disc}}[v] & = N_{\text{disc}}[\rightarrow] = N_{\text{disc}}[\mid] & 3 \text{ disjuncts in the DNF-representation}
\end{array}$$

This accounts for explanandum (IV) of §7.3.5. All of the connective behaviour covered in that paragraph has now been explained.

Notice that the formulae derived in this paragraph only cover the cases with $N_D = 2$, while assuming a restricted use of the connective with only one set of arguments, each of which must contain a different meaning space dimension (as mentioned before in note 43). Also, these formulae represent an absolute number of cases, rather than a proportion. Plotting the fomulae therefore does not yield graphs resembling those of appendix 4. Yet, even though the formulae simplify the reality of the simulation, they have been shown to explain all the connective behaviour observed in §7.3.5.

7.3.7 Connective specificity as a function of positivity / negativity and number of DNF disjuncts

How can these numerical observations be related to real communication? The previous paragraphs have demonstrated that the number of times a particular connective discriminates depends on how many topic objects it can include in its description, and how many background objects it can exclude. Negative connectives such as \mid include a lot of topic objects, but also include a large number of background objects. On the other hand, positive connectives like \wedge exclude a large number of background objects, but also exclude a large number of topic objects. Positive connectives are more *specific*: they rule out more cases. A specific connective works well if a small number of topics needs to be discriminated from a large number of background objects. Negative connectives are less specific and more general, which works well if a large number of topics needs to be picked out and there is only a small number of background objects.

What makes a connective specific or general? There are two factors, namely amount of negation and number of DNF disjuncts.

Connectives that are true in many cases (e.g. v or \mid) are general, while connectives that are true in only one case (e.g. \wedge or BUT) are specific. Amount of negation is the other factor: asserting what colour object *does* have (a positive statement) is more specific than asserting what colour an object

does *not* have, provided that there are more than 2 colours in the universe (i.e. $N_R > 2$). Of the connectives with one DNF-disjunct, \wedge is the connective that includes no negation, BUT has more negation since it makes a positive statement about its first argument and a negative one about its second argument, and \downarrow contains most negation as it makes a negative statement about both its arguments. The other connectives are composed of \wedge , BUT / ALTHOUGH and \downarrow , and have amounts of negation calculable from these four.

Table 3. Order of specificity of all the connectives.

Interestingly, the two connectives that are realized in all natural languages, \wedge and \vee , are the most specific and least specific connective of the positive group, respectively. Another commonly lexicalised connective, \downarrow , is the most specific connective of the negative group.

In run 2, the moderately negative and positive connectives are the most frequent connectives on average. The connectives on the extreme ends of their groups, such as \wedge , are the least frequent connectives. Generally, the less extreme a connective is within its group, the more frequently it is used, although this correlation is not perfect. The reader may verify this from the total frequencies of the connectives in run 2, presented in fig. 3, also repeated below in fig. 8a.

Now, it seems that some of the combinations of N_T and N_B present in run 2 are not very realistic. For instance, humans do not usually try to point out 7 objects at a time to another human, trying to make sure he does not confuse them with 1 background object. It will be instructive to do one more run in which we attempt to make the settings of the simulations as realistic as possible.

Data with realistic settings can be obtained from the results of run 2, by eliminating the simulations with the less realistic settings. We remove the simulations with a high number of topics. The settings for N_B , N_R and N_D remain unchanged. We then end up with data where $N_D = 3$, $N_R \in \{2,3,5,7\}$, $N_T \in \{1,2\}$ and $N_B \in \{1,2,3,5,7\}$. From these data we get the mean total frequencies depicted in fig. 8b below. For comparison the original mean frequencies of run 2 are repeated in fig. 8a.

Figure 8a [= Figure 3]. Mean number of times $N_{disc}[C]$ in which each connective C successfully discriminates topic from background in run 2.

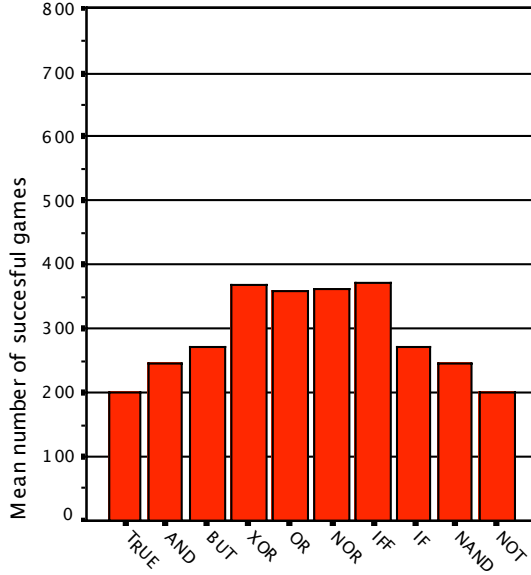
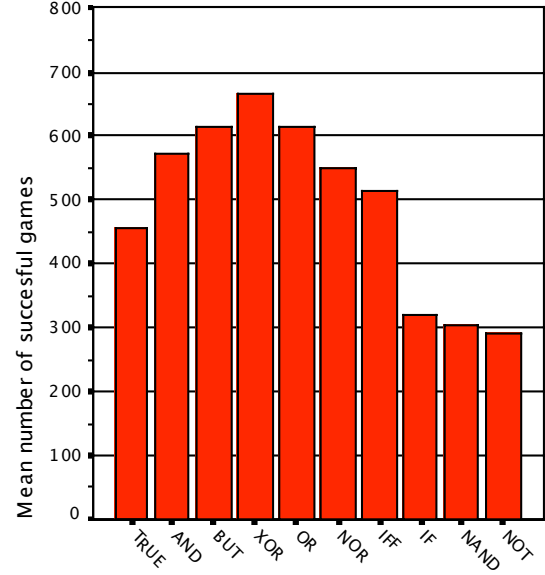


Figure 8b. Mean number of times $N_{disc}[C]$ in which each connective C successfully discriminates topic from background in run 2. Simulations with unrealistic settings eliminated from the run.



As expected, the frequencies of the positive connective are higher in the ‘realistic’ version of run 2, since the simulations in which the negative connectives performed best, those with high N_T , were eliminated. The frequencies in this ‘realistic’ run are still very much unlike the frequencies we would expect from natural language. The most frequent connective is \vee , which is common in natural language. The connective \wedge is also relatively frequent, although outperformed by \vee , $\underline{\vee}$ and BUT and only marginally more frequent than \downarrow or \Leftrightarrow . TRUE and \neg remain infrequently used.

In §7.3.2 it was suggested that a large frequency of the one-placed connectives TRUE and \neg might spontaneously arise, without making the assumption of a preference for one-placed over two-placed connectives. It is clear that this prediction has not borne out. Perhaps introducing such a prespecified preference for one-placed connectives would change the frequencies for the two-placed connectives in a way that better resembles the natural frequency distribution of logical connectives. Building in this preference into this algorithm means that, once a one-placed connective has been found that works in a particular discrimination game, all two-placed connectives are considered not to apply to that game. Setting this preference will thus reduce the number of times each two-placed connective is used. Importantly, it may affect some connectives more than others, changing the relative frequencies of the two-placed connectives.

However, the data shows that the preference indeed reduces the use of the two-placed connectives, but leaves their relative frequencies mainly intact. This can be seen if we compare fig. 8 with fig. 9 below, which contains graphs depicting the mean frequencies of the connectives in reruns of run 2 and the version of run 2 with unrealistic simulations eliminated, but this time with a preference for one-placed connectives in place:

Figure 9a. Mean number of times $N_{\text{disc}}[C]$ in which each connective C successfully discriminates topic from background in run 2. Preference for one-placed over two-placed connectives.

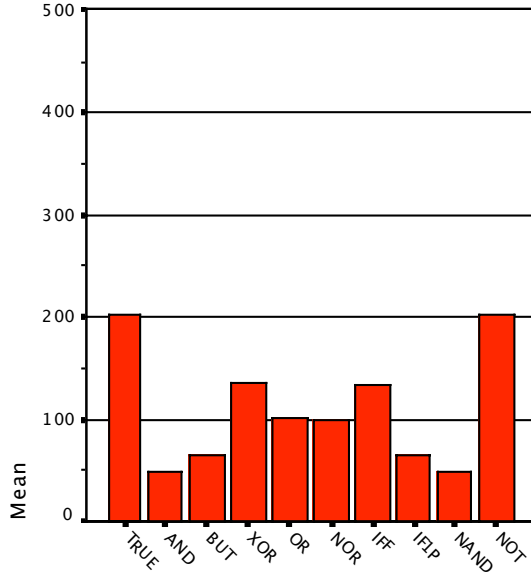
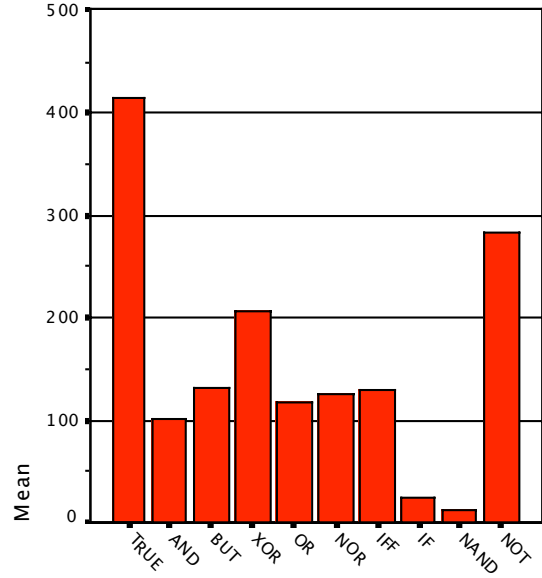


Figure 9b. Mean number of times $N_{\text{disc}}[C]$ in which each connective C successfully discriminates topic from background in run 2. Preference for one-placed over two-placed connectives. Simulations with unrealistic settings eliminated from the run.



We seem to be unable to get a frequency distribution from our simulation that would, if used as input to the Iterated Learning Model, yield holistic phrases for \wedge , \vee and \downarrow and combinatorial realisations for the other connectives.

In Kirby (2001) there were 2 meaning spaces with 5 regions each. Each of these sets of regions had a Zipfian frequency distribution. This means that if the most frequent meaning is used 100 times, the second most frequent is used 50 times, the third 33 times, the fourth 25 times and the fifth 20 times. Of the regions, only the 2 most frequent ones were realized holistically. The frequencies needed to obtain a holophrase for some meaning do depend on the ILM model settings such as bottleneck size (Simon Kirby p.c.). Still, this example should serve to illustrate that in order to get a result with a holophrase for \wedge , \vee and \downarrow , these connectives need not only be the most frequent connectives, but also the most frequent ones by a reasonable margin.

If we cannot obtain such a frequency distribution by trying to construct a realistic simulation based on what we think we know about the real world, perhaps we could reverse-engineer a simulation run that would give us the desired frequency distribution. In order to do that, we should know what combinations of settings work best for the desired connectives \wedge , \vee and \downarrow , and worst for the other connectives.

In table 4 below the connective with the highest frequency is given for each combination of $N_T \in \{1, 2, 3, 5, 7\}$ and $N_B \in \{1, 2, 3, 5, 7\}$. The setting $N_R = 7$ is used throughout the table, because connective frequencies tend to diverge more with higher settings of N_R , making it more likely that a single winner can be found for each combination of N_T and N_B . The results for the other settings of N_R are similar, although there are more cases in which two connectives share the first place in frequency.

Table 4. The most frequent connective for each combination of $N_T \in \{1,2,3,5,7\}$ and $N_B \in \{1,2,3,5,7\}$. $N_R = 7$.

		N_B				
		1	2	3	5	7
N_T	1	\wedge BUT \downarrow TRUE $\neg \leftrightarrow \underline{v}$ $v \rightarrow$	\wedge BUT \downarrow $\leftrightarrow \underline{v}$	\wedge BUT	\wedge	\wedge
	2	$\leftrightarrow \underline{v}$ $v \rightarrow$	\underline{v}	\underline{v}	\underline{v}	\underline{v}
	3	$\leftrightarrow \rightarrow$	\leftrightarrow	\underline{v}	\underline{v}	\underline{v}
	5		\leftrightarrow	\leftrightarrow	v	$\underline{v} v$
	7		\leftrightarrow	\leftrightarrow	\downarrow	$\underline{v} v$

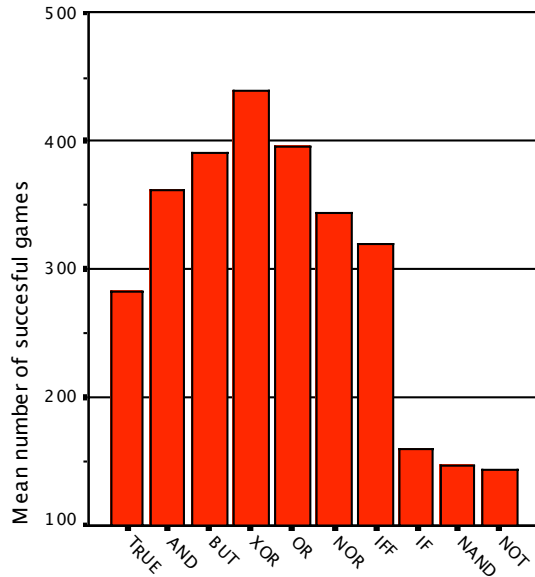
The table confirms our observations about the characteristics of the positive and negative connectives. The first column in table 4 contains the cases with just one background object. As we move down in this column, the number of topic objects increases and the negative connectives keep doing better, while the positive connectives disappear. Finally, once $N_T \geq 5$ and $N_B = 1$, the most extreme connective of the negative group, \downarrow , surfaces as the most commonly used connective. Conversely, the first row contains all the cases with just one topic. As N_B increases, the negative connectives disappear from this row and the positive ones win out. Ultimately, once $N_B \geq 5$, the most extreme connective of the positive group, \wedge , ends up as the winner.

We already knew that negative connectives become ineffective quickly as N_B increases, and that positive connectives become ineffective as N_T increases. So what if both N_B and N_T are high? As can be seen from the table 4, the moderate connectives of each group work best, since they are the best available compromise between the characteristics of both groups. These connectives are the moderately positive and negative connectives \downarrow , v , \leftrightarrow and \underline{v} , the latter two of which are the most common. Once we have seen table 4 it no longer comes as a surprise that these connective were the most frequent ones overall in run 2. Extremely positive connectives win when $N_T = 1$, extremely negative connectives when $N_B = 1$, but the moderate connectives, whether they are positive or negative, win in all the other cases.

Of the four moderate connectives, the negative ones \leftrightarrow and \downarrow occupy the lower left of the table that contains the cases with higher N_T and lower N_B . The positive connectives \underline{v} and v are in the higher right of the table, which contains the cases with higher N_B and lower N_T . Judging from this table, a simulation run that would yield high frequencies for \wedge and v would contain a large number of cases with $N_T = 1$, which is when \wedge works best, and a large number of cases where $N_T > 1$ and $N_B \geq N_T$, which is when \underline{v} and v works best. This seems like a fairly realistic assumption about real linguistic pointing behaviour, although I do not have any independent evidence to back up this judgement.

Eliminating all the simulations with $N_B < N_T$ from run 2, then, reveals the following frequency distribution:

Figure 10. Mean number of times each connective is successfully used in run 2. Simulations with $N_T > N_B$ eliminated.



Indeed, \vee and $\underline{\vee}$ are most frequent connectives. BUT is more frequent than \wedge , though. This reflects the relative order of BUT and \wedge in the cases in which \vee or some other connective is the most frequent connective; BUT apparently does better than \wedge in those cases. Extending the simulation to include more cases with more background objects ($N_B > 7$), an extension that seems realistic, would surely change the relative frequency of BUT and \wedge in favour of the latter, since \wedge , being more specific than BUT, benefits more from more background objects

7.4 Specificity in real world human communication

The main discovery we have made in our simulation is a hierarchy of specificity among the connectives. That saying *P and Q* is more specific than saying *P or Q* has of course for long been common knowledge in logic, semantics and pragmatics. What I have added in the previous paragraphs is a quantification of the notion of specificity for logical connectives.

The formalisation has shown that the specificity of a connective depends on three factors. Two of these, amount of negation and number of DNF-disjuncts, are properties of the connectives themselves. The other factor, number of regions in the meaning space dimension(s) used as an argument to the connective, is a property of the environment. This property of the environment in turn influences the effect amount of negation has on the specificity of the connective.

Since Grice (1975), specificity or informativeness has been taken to be one of the main factors determining the use of language in conversation. Grice thought of communication as a co-operative effort towards a mutual goal. Given this, he felt it would be reasonable to assume there must be rules (*maxims* in Grice's parlance) to which participants in a conversation are expected to conform.

The mutual goal of the conversation partners in Sierra-Santibañez's simulation is to point out the topic objects and set them apart from the background objects: nothing more, nothing less. Achieving this goal is aided by sticking to the following of Grice's (1975: 45–46) maxims:

Maxims of Quantity

1. Make your contribution as informative as required
2. Do not make your contribution more informative than required

Maxims of Manner

1. (...)
2. Avoid ambiguity
3. (...)

The agents conform to the manner of Quantity: the connective they use is as specific as is required to set the topic apart from the background, but not more specific than that. An expression with the right amount of informativeness has to be an *unambiguous* expression. If the expression were ambiguous, it would also pick out some other set of objects that are not in the topic: the expression would thus not be specific enough to achieve the communicative goal. Obeying the maxim of Manner and obeying the first maxim of Quantity amount to the same thing in this simulation.

I have found that different situations call for different amounts of specificity or informativeness in order to achieve this communicative goal: sometimes an unspecific connective such as *v* or even *|* suffices. Importantly, the kinds of ‘pointing-out situations’ that humans are most likely to find themselves in, are the situations to which the connectives *∧* and *v* / *∨* are best suited.⁴⁴ The former works best when only one object needs to be pointed out, irrespective of the number of background objects. The latter works best in the other cases where the number of topics is larger than one, but not larger than the number of background objects.

Obviously the goals of human communicators extend beyond the mere pointing out of objects by referring to their perceptual properties, and connectives are used in more contexts than that.

First of all, the statements humans make in real life are not usually confined to a small universe consisting of just a few objects. Instead, they are made with the world as background, or just a loosely defined ‘context of the conversation’. If the goal of the speaker is still assumed to be the unique identification of worlds or objects, we may expect a frequent use of the specific connectives. Also, many real-world meaning space dimensions can be divided into vast or possibly infinite numbers of regions. Take the property of walking, for instance. We can take walking to be a region in a meaning space dimension [ACTIONS THAT CAN BE TAKEN], with an infinite number of regions. Then, *the guy that is walking* may refer to thousands of persons, but *the guy that is walking and singing* to millions of possibly billions. Similarly, *John walks* may be true in just one world, but *John does not walk* is true all the possible world in the Universe, minus that one world in which John does walk. Seen from this perspective, the negative connectives are uninformative almost to the point of uselessness.

Still, fairly uninformative connectives such *or* and *nor* are still used regularly in human language.⁴⁵ One reason is that humans have another maxim to conform to, one that is not an issue for the simulation agents:

Maxims of Quality (Grice 1975: 46)

1. (...)
2. Do not say for which you lack adequate evidence.

A speaker may refer to the *red or the green ball* simply because he does not know whether the ball is red or green. The artificial agents by contrast always have all the existing information about the objects at their disposal.

⁴⁴ Both *v* and *∨* occurred in these situations. *∨* seems a bit more frequent, but I would not dare to derive any conclusions from this about whether English *or* actually means *v* or *∨* (see §3.3.1 for details on this debate).

⁴⁵ We should keep in mind though that *and* is by far the most frequent connective in English: corpus-based research shows that it is about 6 times more frequent than *or* is (Ohori 2004: 61).

Also, the maxims of Quantity and Ambiguity may apply differently in the case of humans. First of all, complete unambiguity is never needed in real communication as humans have a large amount of background knowledge they can use to aid interpretation of an utterance. The agents in the simulation on the other hand are completely ignorant of everything in the world around them except what they see in their sensors. Also, humans may consciously break Grice's conversational maxims if they do not want to achieve the goal of accurate communication: they may choose to be deliberately ambiguous or uninformative in order to manipulate their partner, out of laziness, etc.

Summarising, we seen have that real human communication is or may be different from the simulation in all its aspects. The communicative goals may be different, the world is infinitely more complex, the agents are more complex and have much more background knowledge and the rest of the language is more complex and allows for conjunction of many categories.

Extending the simulation to the real world is thus very difficult and this leaves us with the question what it is we have actually proved. The question of external validity is an issue for all experiments in social science (e.g. Porte 2002: 45–47), and also for all computer simulations of complex phenomena. A famous example of the latter are Artificial Life simulations, whose applicability to biological reality is often questioned (e.g. Miller 1995). Recently, Bickerton (In Press) has expressed similar concerns about language evolution simulations specifically. Luc Steels himself, the founding father of the Language Games paradigm to which the Sierra-Santibañez simulation belongs, explicitly refuses to make any claim of empirical validity of his simulations to humans or animals (e.g. Steels 1996: 339). He merely aims to explore the theoretical possibilities of the evolution of meaning.

Ultimately then, we should limit ourselves to concluding that the simulation has demonstrated the beneficiality of observing a subset of Grice's maxims to achieving a certain limited goal, namely discriminating topics from background objects, and that in turn what connective best observes Grice's maxims depends on how many objects need to be discriminated from how large a background. Tantalisingly, we notice that in a number of situations that are, within the limited bounds of the reality of the simulation, the ones most reminiscent of human communicative situations, Grice's maxims are best observed by use of the connectives \wedge or \vee . However, the value of this result depends wholly on the somewhat indeterminable equivalence of the model to reality.

8 The nature of the compositionality of connectives

A crucial part in the simulation run by Kirby (2001) is the representation of the meanings that are to be lexicalised. This meaning representation is the yardstick we use to decide whether some signal is holistic or compositional: it is compositional if it is composed of several elements corresponding to different elements of the meaning representation; it is holistic if it collapses several elements of the meaning into one signal.

This problem is another tough nut to crack if we wish to apply the Iterated Learning Model to the problem of holistic and compositional logical connectives in human language. Unfortunately, all I am able to do in this chapter is describe why the model of Kirby (2001) cannot properly represent the particular kind of compositionality displayed by the connectives. The rest of the chapter is a slightly speculative exploration of what kinds of changes could be made to the model so it *could* do the trick.

The meaning representation used in Kirby (2001) is a bit vector. Bit vector meanings are defined as feature vectors representing points in a meaning space (cf. the description of Brighton, Smith and Kirby 2005: 185–186). A meaning space is defined by a number N_D of dimensions, each of which is divided into a number N_R of regions. A meaning space thus defined is filled with a set of N_R^{ND} possible points. For instance, a meaning space M specified by $N_D = 2$ and $N_R = 2$ for each dimension would be represented by the set $M = \{(a,a),(a,b),(b,a),(b,b)\}$ of meaning space points.

The meaning representation described here is similar to the representation Sierra-Santibañez uses for objects in her (2001) simulation. In this simulation, a meaning space over objects is used, each point in the meaning space representing a possible object. The dimensions in the meaning space are the perceptual features of the objects, e.g. COLOUR, SIZE or HEIGHT. The regions represent the values that each of these features can accommodate. For instance, the meaning dimension of size could take three values: LARGE, MEDIUM and SMALL.

For the representation of the logical connectives in a bit vector, an obvious strategy *prima facie* would be to construct a meaning space M_{con} based on their truth tables, as follows:

$N_D = 4$; $N_R = 2$ for each dimension D

$D_1 \in (1,0)$	$D_1 = 1$ if the connective yields true if antecedent and consequent are both true $D_1 = 0$ if the connective yields false if antecedent and consequent are both true	
$D_2 \in (1,0)$	$D_2 = 1$ if the connective yields true if antecedent is true and consequent is false $D_2 = 0$ if the connective yields false if antecedent is true and consequent is false	
$D_3 \in (1,0)$	$D_3 = 1$ if the connective yields true if antecedent is false and consequent is true $D_3 = 0$ if the connective yields false if antecedent is false and consequent is true	
$D_4 \in (1,0)$	$D_4 = 1$ if the connective yields true if antecedent and consequent are both false $D_4 = 0$ if the connective yields false if antecedent and consequent are both false	

Set of possible meaning space points M_{con}

{	(0,0,0,0) ‘NEVER’	(0,1,0,0) ‘BUT’	(1,0,0,0) ‘ \wedge ’	(1,1,0,0) ‘AM’
	(0,0,0,1) ‘ \downarrow ’	(0,1,0,1) ‘MN’	(1,0,0,1) ‘ \leftrightarrow ’	(1,1,0,1) ‘ \leftarrow ’
	(0,0,1,0) ‘ALTHOUGH’	(0,1,1,0) ‘ \vee ’	(1,0,1,0) ‘MA’	(1,1,1,0) ‘v’
	(0,0,1,1) ‘NM’	(0,1,1,1) ‘ \uparrow ’	(1,0,1,1) ‘ \rightarrow ’	(1,1,1,1) ‘ALWAYS’

In a system like this, we could get a holistic representation for an entire meaning like (1,0,0,0) ‘ \wedge ’. Such a holophrase could be *and*, like in English. The compositional connectives would be completely

unlike those in English, though. They may consist of up to four meaning particles, each corresponding to one or more of the meanings features D_x . For instance, if a lexeme *ka* had evolved to represent [$D_1 = 0 \wedge D_2 = 1$], another lexeme *poom* existed meaning [$D_3 = 1$] and finally *ba* means [$D_4 = 0$], then a compositional representation for (0,1,1,0) ‘ \vee ’ might be *ka-poom-ba*.

The bit vector representation in Kirby (2001) appears inspired by word morphology. The derivation of compositional connective representations in logic (mirroring language to an extent, see §3.4) is more like syntax than like morphology. Each of the connectives is a function, so it needs to be in a particular syntactic juxtaposition with its arguments: we cannot take the propositions out of the equation if we want to describe compositional connective representations. Also, the syntax is recursive: connectives can be applied to each other so that sentences of arbitrary length can be constructed. Bit vectors, on the other hand, are of finite length and their meaning components come in a fixed order.

It may seem that we could just try and expand the meaning representation used in Kirby (2001). Other studies into the emergence of compositionality have been done using predicate logic (De Beule and Bergen 2006; Kirby 2000), some successfully employing recursive predicates (Kirby 2002). The findings about regular and irregular lexicalisations from the (2001) study could be replicated with these more complex meaning representations (Simon Kirby p.c.).

Although we cannot compositionally represent connectives in this syntactic manner without including the propositions, we may be able to abstract away from that problem by representing including variable proposition placeholders P and Q in the connective meanings. We could thus assume that, say, \leftrightarrow is represented in our minds as $\lambda P \lambda Q [(P \wedge Q) \vee (\neg P \wedge \neg Q)]$.

It is obvious that this would be putting the cart before the horse. The very reason that a representation like $\lambda P \lambda Q [(P \wedge Q) \vee (\neg P \wedge \neg Q)]$ can be said to be equivalent to \leftrightarrow are the functions \wedge , \vee and \neg in the representation. We still have not represented the meanings of those functions.

The functions of predicates are noticeably absent in all simulations of compositionality I have discussed. These simulations assume that intensional meanings, the predicate-argument structures themselves, suffice as a meaning representation. The need has not yet been felt to enrich these predicates with function to compute the ‘true meaning’—the Fregean extension—of predicate-argument structures.

We *do* need to represent these functions with the connectives. What allows all the 16 connectives to be compositionally represented with syntactic derivations is the fact that the set of their 16 functions has the special property that applying one set member to another one will always result in the meaning of another set member, provided that the same arguments are used every time. This is unlike all other predicates in language, the application of whose functions results in a new meaning each time, *ad infinitum*.

The best representation we have for connective functions are the various incarnations of the truth table. We have seen above that this truth table is in no way morphologically mirrored in the words for the connectives in existing natural languages. Instead we have holistic words for a few base connectives (*and* ‘ \wedge ’, *or* ‘ \vee ’, *not* ‘ \neg ’ and *nor* ‘ \downarrow ’), which are applied to each other to create other connective functions⁴⁶. Ultimately, it looks like the synthetic ‘Bickertonian’ scenario for the evolution of compositionality is more appropriate in the case of the connectives than an analytic one: words

⁴⁶ *Nor* is taken by many (e.g. Gazdar & Pullum 1976; Jaspers 2005: 54) to morphologically reflect compositional meaning. (This morphological transparency is disputable, though, see §4.2 for discussion). The hypothesised meaning elements, are *not* and *or*, which are normally taken to correspond to the full connective meaning space points \neg and \vee . So, *not* and *or* are not ‘parts’ of the connective meaning of \downarrow in the sense of representing a subset of the meaning features that \downarrow is composed of (i.e. $D_1 = 0$, $D_2 = 0$, $D_3 = 0$, $D_4 = 1$). Instead, *nor* looks more like it was originally derived in the syntactic way, through iterated application of the functions of *not* and *or*. This syntactically derived word group *not + or* could later have fossilised into *nor*.

came first and these words were used in compositional representations later (cf. ch. 5).

The question remains: why are some of the connectives used as base connectives, and why are other connectives derived by iterated application of these base connectives?

No computer simulations exist that model the Bickertonian scenario for the development of connectivity. In a simulation that could somehow simulate the emergence of connective derivations by recursive application, we would need to expand not the semantic representations of the original model, but rather its learning mechanism. The learning mechanism in the original simulation discovered compositional rules by looking for correlations between bits of meaning and bits of holophrases. The learning mechanism in a simulation producing connective combinations would need more ‘creative power’ and learn how to apply connective functions to each other.

There are numerous issues here, of course. Will the bottleneck principle, that led to compositionality and frequency effect in the first place, still work if we change the learning mechanism this radically? What would a learning mechanism look like that somehow ‘sees’ that it can apply the connectives to each other? I cannot begin to answer these questions here.

I do remain confident that frequency would influence which connectives are used as base connectives, and which are derived by iterated application of these base connectives. A possible motivation for such an assumption could be the fact that combinatorial lexicalisations of connectives are, by definition, longer than holistic ones: a well-conformed observation by Zipf (1935) is that longer phrases tend to be less frequent. A system employing the most frequent connectives as base connectives would be optimal in this sense.

A final concern may be that even if we could evolve connective combination through function application, that would just assume that language syntax combines connectives like logic, ignoring the fact that language has its own idiosyncratic ways of combining connectives, too (cf. §3.4). That said, the current utterances produced by the ILM models—e.g. *s*, *yuqeg* or *uhlbg* in Kirby (2001)—do not look very much like natural language yet, either. Although I would like to hold myself to higher standards than the rest of the field here, I need to be realistic and defer the task of developing simulations that yield full-fledged human language to future generations of researchers.

9 Conclusions

The objective of this thesis was to find a general solution to an old chestnut of semantics: the absence in natural language of single-morpheme lexicalisations for many possible logical connectives, which are lexicalised instead as compositions of other connectives. From an investigation of currently available accounts we concluded that a cultural evolutionary approach is needed, specifically one that pays sufficient attention to the compositional structure of language.

Such an approach was found in the computational model of the emergence of compositionality by Kirby (2000; 2001). This model demonstrates how, as a consequence of the transmission of language across generations of learners, frequent meanings evolve to be represented as irregular holistic phrases, whereas infrequent meanings are coupled with compositional representations. I have tried to apply this model to the research problem and thus hypothesised that (a) the connectives differ with respect to the frequency with which each of them needs to be expressed and (b) that the concept of regular and irregular lexicalisations from Kirby's model can be applied to the particular kind of compositionality that is found in the representation of connectives.

It proved a considerable challenge finding the right evidence for both these hypotheses. As for hypothesis (a), I could conceive of two factors that might give rise to a difference in frequency between the connectives. A connective may be used less, because human minds find it hard to reason with, or because it is less useful for communication. The first explanation remained unsubstantiated. Psychological theories of reasoning difficulty provided no obvious path to frequency data, while reasoning difficulty by itself was shown to make empirically false predictions about which connectives should be present in language.

I have tried to substantiate the second explanation by using a simplified model of human communication, since there are obvious difficulties with extracting such frequency data from the real world. The model, Sierra-Santibañez (2001), was limited to what we might call the *demonstrative* function of language. It consisted of agents aiming to discriminate sets of topic objects from background objects. For this purpose the agents used descriptions involving perceptual properties of the objects, conjoined by the connectives.

Even this limited communication model contained many variables influencing the frequencies of the connectives, such as number of topic objects, number of regions per meaning space dimension, etc. It was not immediately clear what setting of these variables would simulate natural communication best. This uncertainty called for an exploratory analysis of the effects of all the variables and their interaction effects.

The analysis of the simulation results revealed that some connectives thrived in a situation with many topic objects, while others fared comparatively better when there were many background objects. A hierarchy was discovered among the connectives with respect to the property of their response to changes in these two variables. The property itself was identified as *specificity*, a notion familiar from Gricean pragmatics.

Adherence to Grice's maxim of Quantity—'be as specific as you need to be'—is considered to be a prerequisite for successful communication and is shown to be a condition for success in our simulation, too. Due to the particular position of \wedge on the specificity hierarchy, adherence to the maxim of Quantity is best served by that connective in the case where there is one topic object and any number of background objects. On the other hand \vee or $\underline{\vee}$ work best in cases where the number of topic objects is larger than one, but not larger than the number of background objects.

These two simulations settings seem to be ones most reminiscent of human communicative situations—within the limited bounds of the reality of the simulation, that is. \wedge and \vee are the connectives most frequently used by the agents in those human-like situations, because they are the ones that best conform to the maxim of Quantity. We must be careful with extending the findings

from this model to the real world, though, because there is a lack of independent evidence backing up the external validity of this model for real world communication.

Now that we had found some evidence that \wedge and $\vee / \underline{\vee}$ may in fact be the most frequent among the 16 connectives, I have tried to concretely apply Kirby's model of the link between frequency and compositionality to the case of the connectives. However, it proved difficult to implement the particular way in which language constructs compositional lexicalisations of connectives. The learning mechanism in Kirby's model develops compositional representations that are like word morphology, with a fixed order of meaning components and fixed length. However, connective combination in language is more like syntax: some connectives are used as base connectives, from which other connectives are derived by recursive application of their functions. In order to adapt the model to develop such representations we would have to radically change the learning mechanisms, up to a point where a lot of the basic tenets of the model—ones that cause the development of compositionality in the first place—might have to be altered. Investigating these problems was unfortunately not possible within the scope of a master's thesis.

I do remain convinced that frequency has a very important role in explaining why only some connectives have single-morpheme lexicalisations, while others are derived syntactically from those connectives. The relationship between frequency and compositionality is well attested in natural language (e.g. Pinker 1999: 123–125) and the causes for its emergence have been demonstrated in Kirby's Iterated Learning Model (Kirby 2000; 2001; 2002). Also, Zipf's (1935) observation that shorter words are more frequent than longer ones has been confirmed by empirical evidence (e.g. Sigurd, Eeg-Olofson and Van de Weijer 2004) and has plausible causes in the principle of *least effort* (Zipf 1949; cf. also Vogt 2004). By definition, syntactical derivations of base connectives are longer and more complex than those base connectives themselves. A principle of least effort would surely favour a system in which the most frequently needed connectives are realized as single morphemes from which all others are derived. Such a system would allow the language user to use the shortest and simplest expressions on average. The evidence from our communication simulation suggests that those most frequent connectives might very well be \wedge and \vee or $\underline{\vee}$.

Where should research go from here? Of course there are the numerous issues with my account that should be resolved. There are important concerns about the realism and empirical validation of the model of communication: we need new models that are based on independent evidence on how, in what circumstances and with which goals human communication is conducted. In particular, more recent and sophisticated models of pragmatics than Grice's (e.g. Sperber and Wilson 1995) should be included. We also need to develop a credible mechanism that could account for the way in which humans learn to represent connectives compositionality as a structure of repeated mutual application of connectives, possibly within the framework of the pre-existing models, but possibly not.

Finally, perhaps research should also reconsider the properties of human mind. In this thesis I have proposed an explanation for the problem of connectives in human language that wholly separates that problem from the matter of 'how humans reason'. Instead I have relied on environmental factors and a general need for communication as an explanation. On one hand it may seem that ignoring human minds is taking an important factor out of the equation—the limited research I did do on the possible connection between human reasoning and language connectives may have been insufficient. On the other hand, we should realize that possible innate preferences for connectives and human reasoning abilities themselves must be a result of environmental pressures and possibly a desire for communication. Instead of ignoring the facts of human reasoning, this thesis may well be taken to suggest a direction in which explanations for the evolution of those abilities might be found.

References

- Barett, Robert & Alfred Stenner. (1971). The myth of the exclusive 'or'. *Mind* 80 (317): 116–121.
- Bar-Lev, Zev & Arthur Palacas. (1980). Semantic command over pragmatic priority. *Lingua* 51 (2–3): 137–146.
- Barnhart, Robert, (ed.). (1988). *The Barnhart dictionary of etymology*. Bronx, NY: Wilson.
- Batali, John. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In: Ted Briscoe (ed.), *Linguistic evolution through language acquisition* (pp. 111–172). Cambridge: Cambridge University Press.
- Benz, Anton, Gerhard Jäger & Robert van Rooij. (2006). An introduction to game theory for linguists. In: Anton Benz & Gerhard Jäger (eds.), *Game theory and pragmatics* (pp. 1–82). Basingstoke, UK: Palgrave Macmillan.
- de Beule, Joachim & Benjamin Bergen. (2006). On the emergence of compositionality. In: Angelo Cangelosi, Andrew Smith & Kenny Smith (eds.), *The evolution of language: proceedings of the 6th International Conference (EVOLANG6), Rome, Italy, 12–15 april 2006* (pp. 35–42). Hackensack, NJ: World Scientific.
- Bickerton, Derek. (1995). *Language and human behaviour*. Seattle: University of Washington Press.
- Bickerton, Derek. (2000). How protolanguage became language. In: Chris Knight, Michael Studdert-Kennedy & James Hurford (eds.), *The evolutionary emergence of language: social function and the origins of human form* (pp. 264–284). Cambridge: Cambridge University Press.
- Bickerton, Derek. (In press). Language evolution: a brief guide for linguists. *Lingua*.
- Braine, Martin & David O'Brien. (1998). How to investigate mental logic and the syntax of thought. In: Martin Braine & David O'Brien (eds.), *Mental logic* (pp. 45–61). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brighton, Henry, Kenny Smith and Simon Kirby. (2005). Language as an evolutionary system. *Physics of Life Reviews* 2 (3): 177–226.
- Bybee, Joan. (2003). Mechanisms of change in grammaticalization: the role of frequency. In: Brian Joseph & Richard Janda (eds.), *The handbook of historical linguistics* (pp. 602–622). Malden, MA: Blackwell.
- Carnap, Rudolf. (1956). *Meaning and necessity: a study in semantics and modal logic*. Second, enlarged edition. Chicago: University of Chicago Press. First edition published 1947.
- Carstairs-McCarthy, Andrew. (1999). *The origins of complex language: an enquiry into the evolutionary beginnings of sentences, syllables and truth*. Oxford: Oxford University Press.
- Chang, Chin-Liang & Richard Char-Tung Lee. (1973). *Symbolic logic and mechanical theorem proving*. New York: Academic Press.
- Chierchia, Gennaro & Sally McConnell-Ginet. (2000). *Meaning and grammar: an introduction to semantics*. Second edition. Cambridge, MA: MIT Press.
- Christiansen, Morten & Joseph Devlin. (1997). Recursive inconsistencies are hard to learn: a connectionist perspective on universal word order correlations. In: Michael Shafto & Pat Langley, *Proceedings of the 19th annual conference of the Cognitive Science Society, August 7–10, 1997, Stanford University, Stanford, CA* (pp. 113–118). Mahwah, NJ: Lawrence Erlbaum Associates.
- Christiansen, Morten & Simon Kirby. (2003). Language evolution: the hardest problem in science? In: Morten Christiansen & Simon Kirby (eds.), *Language Evolution* (pp. 1–15). Oxford: Oxford University Press.
- Collinson, William. (1937). *Indication: a study of demonstratives, articles and other 'indicators'*. [Language Monographs 17]. Baltimore, MD: Waverley Press.
- Deacon, Terrence. (1997). *The symbolic species: the co-evolution of language and the brain*. New York: W.W. Norton & Co.
- Dieterich, Thomas & Donna Napoli. (1982). Comparative 'rather'. *Journal of Linguistics* 18 (1): 137–165.
- Du Bois, John. (1985). Competing motivations. In: John Haiman (ed.), *Iconicity in syntax*:

- proceedings of a symposium on iconicity of syntax, Stanford, June 24–6, 1983* (pp. 343–365). Amsterdam: John Benjamin.
- Dunbar, Robin. (1998). *Grooming, gossip and the evolution of language*. Cambridge, MA: Harvard University Press.
- Forbes, Graeme. (1994). *Modern logic: A text in elementary symbolic logic*. New York: Oxford University Press.
- Forrest, Peter. (2006). The identity of indiscernibles. In: Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2006 edition. Stanford, CA: Stanford University. Available online at <http://plato.stanford.edu/archives/sum2006/entries/identity-indiscernible/>
- Gazdar, Gerald & Geoffrey Pullum. (1976). Truth-functional connectives in natural language. In: Salikoko Mufwene, Carol Walker & Sanford Steever (eds.), *Papers from the twelfth regional meeting, Chicago Linguistic Society, April 23–25, 1976* (pp. 220–234). Chicago: Chicago Linguistic Society.
- Gil, David. (1991). Aristotle goes to Arizona and finds a language without *and*. In: Dietmar Zaefferer (ed.), *Semantic universals and universal semantics* (pp. 96–130). [Groningen-Amsterdam Studies in Semantics 12]. Berlin: Foris.
- Grice, Herbert Paul. (1975). Logic and conversation. In: Peter Cole & Jerry Morgan (eds.), *Speech acts* (pp. 41–58). [Syntax and Semantics 3]. New York: Academic Press.
- Haiman, John. (1985). *Natural syntax: iconicity and erosion*. [Cambridge Studies in Linguistics 44]. Cambridge: Cambridge University Press.
- Hawkins, John. (1994). *A performancy theory of order and constituency*. Cambridge: Cambridge University Press.
- Hellemans, Ludo. (2003). Luc Steels: het verschil tussen mens en machine vervaagt. *Intermediair* Dec. 11: 24–29.
- Horn, Laurence. (1989). *A natural history of negation*. Chicago: University of Chicago Press.
- Hurford, James. (1974). Exclusive or inclusive disjunction. *Foundations of Language* 11: 409–411.
- Hurford, James. (1990). Nativist and functional explanations in language acquisition. In: Iggy Roca (ed.), *Logical issues in language acquisition* (pp. 85–136). [Linguistic Models 15]. Dordrecht, NL: Foris.
- Hurford, James. (2003). Why synonymy is rare: fitness is in the speaker. In: Wolfgang Banzhaf, Thomas Christaller, Peter Dittrich, Jan Kim & Jens Ziegler (eds.), *Advances in artificial life: 7th European Conference, ECAL 2003, Dortmund, Germany, September 14–17, 2003, Proceedings* (pp. 442–451). Berlin: Springer, pp.
- Hurford, James. (In press). The origin of noun phrases: reference, truth and communication. *Lingua*.
- Jaspers, Dany. (2005). *Operators in the lexicon: on the negative logic of natural language*. Ph.D. dissertation, Universiteit Leiden. Utrecht: LOT.
- Johnson-Laird, Philip. (1999). Deductive reasoning. *Annual Review of Psychology* 50: 109–135.
- Johnson-Laird, Philip. (2001). Mental models and deduction. *Trends in Cognitive Sciences* 5 (10): 434–442.
- Johnson-Laird, Philip & Ruth Byrne. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates.
- Johnson-Laird, Philip, Ruth Byrne & Walter Schaeken. (1992). Propositional reasoning by model. *Psychological Review* 99 (3): 418–439.
- Keenan, Edward & Leonard Faltz. (1985). *Boolean semantics for natural language*. Dordrecht, NL: D. Reidel.
- Kirby, Simon. (1999). *Function, selection, and innateness: the emergence of language universals*. Oxford: Oxford University Press.
- Kirby, Simon. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In: Chris Knight (ed.), *The evolutionary emergence of language: social function and the origins of linguistic form* (pp. 303–323). Cambridge: Cambridge University Press.

- Kirby, Simon. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation* 5 (2): 102–110.
- Kirby, Simon. (2002). Learning, bottleneck and the evolution of recursive syntax. In: Ted Briscoe (ed.), *Linguistic evolution through language acquisition* (pp. 173–203). Cambridge: Cambridge University Press.
- Kirby, Simon & James Hurford. (2002). The emergence of linguistic structure: an overview of the Iterated Learning Model. In: Angelo Cangelosi & Domenico Parisi (eds.), *Simulating the evolution of language* (pp. 121–148). London: Springer Verlag.
- Kirby, Simon, Kenny Smith & Henry Brighton. (2004). From UG to universals: linguistic adaptation through iterated learning. *Studies in Language* 28 (3): 587–606.
- Lewis, David. (1969). *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.
- Li, Wentian. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory* 38 (6): 1842–1845.
- Lightfoot, David. (1991). Subjacency and sex. *Language & Communication* 11 (1–2): 67–69.
- Link, Godehard. (1998). *Algebraic semantics in language and philosophy*. [CSLI Lecture Notes 74]. Stanford, CA: CSLI Publications.
- Maynard-Smith, John. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.
- Miller, Geoffrey. (1995). *Artificial Life as theoretical biology: how to do real science with computer simulation*. [Cognitive Science Research Paper 378]. Brighton, UK: University of Sussex, School of Cognitive and Computing Sciences.
- Newmeyer, Frederick. (1991). Functional explanation in linguistics and the origins of language. *Language & Communication* 11 (1–2): 3–28.
- Newmeyer, Frederick. (1998). *Language form and language function*. Cambridge, MA: MIT Press.
- Ohuri, Toshio. (2004). Coordination in mentalese. In: Martin Haspelmath (ed.), *Coordinating constructions* (pp. 41–66). [Typological Studies in Language 58]. Amsterdam: John Benjamins.
- Peirce, Charles. (1933). A Boolean algebra with one constant. In: Charles Hartsthorpe & Paul Weiss (eds.), *Collected papers of Charles Peirce*. Volume IV: *The simplest mathematics* (pp. 13–18). Cambridge, MA: Harvard University Press.
- Pinker, Steven. (1999). *Words and rules: the ingredients of language*. London: Weidenfeld & Nicolson.
- Pinker, Steven & Paul Bloom. (1990). Natural language and natural selection. *The Behavioral and Brain Sciences* 13 (4): 707–784.
- Porte, Graeme. (2002). *Appraising research in second language acquisition: a practical approach to critical analysis of quantitative research*. [Language Learning and Language Teaching 3]. Amsterdam: John Benjamins.
- Quine, Willard Van Orman. (1952). *Methods of logic*. London: Routledge & Kegan Paul.
- Quine, Willard Van Orman. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Rader, Aaron & Vladimir Sloutsky. (2001). Conjunctive bias in memory representations of logical connectives. *Memory & Cognition* 29 (6): 838–849.
- Rips, Lance. (1994). *The psychology of proof: deductive reasoning in human thinking*. Cambridge, MA: MIT Press.
- van Rooij, Robert. (2004). Evolution of conventional meaning and conversational principles. *Synthese* 139 (2): 331–366.
- van Rooij, Robert. (2005). Evolutionary motivations for semantic universals. Unpublished manuscript, Universiteit van Amsterdam.
- van Rooij, Robert. (2006). Evolutionary games and semantic universals. In: Angelo Cangelosi, Andrew Smith & Kenny Smith (eds.), *The evolution of language: proceedings of the 6th International Conference (EVOLANG6), Rome, Italy, 12–15 april 2006* (pp. 356–363). Hackensack, NJ: World Scientific.

- Schmerling, Susan. (1975). Asymmetric conjunction and rules of conversation. In: Peter Cole & Jerry Morgan (eds.), *Speech acts* (pp. 211–232). [Syntax and Semantics 3]. New York: Academic Press.
- Schumm, George & Stewart Shapiro. (1990). Expressive completeness and decidability. *Notre Dame Journal of Formal Logic* 31 (3): 576–579.
- Seuren, Pieter. (1985). *Discourse semantics*. Oxford: Blackwell.
- Seuren, Pieter. (1998). *Western linguistics: an historical introduction*. Oxford: Blackwell.
- Sheffer, Henry. (1913). A set of five independent postulates for Boolean algebras, with application to logical constants. *Transactions of the American Mathematical Society* 14 (4): 481–488.
- Sierra–Santibañez, Josefina. (2001). Grounded models as a basis for intuitive reasoning: the origins of logical categories. In: Silvia Coradeschi & Alessandro Saffiotti (eds.), *Anchoring symbols to sensor data in single and multiple robot systems: papers from the 2001 AAAI Fall Symposium, November 2–4, North Falmouth, Massachusetts* (pp. 101–108). Menlo Park, CA: AAAI Press.
- Sigurd, Bengt, Mats Eeg-Olofson & Joost van de Weijer. (2004). Word length, sentence length and frequency: Zipf revisited. *Studia Linguistica* 58 (1): 37–52.
- Sloutsky, Vladimir & Yevgeniya Goldvarg. (2004). Mental representation of logical connectives. *Quarterly Journal of Experimental Psychology* 57a (4): 636–665.
- Sperber, Dan & Deirdre Wilson. (1995). *Relevance: communication and cognition*. Second edition. Oxford: Blackwell. First edition published 1986.
- Steels, Luc. (1996). Perceptually grounded meaning creation. In: Mario Tokoro (ed.), *ICMAS-96: proceedings. Second International Conference on Multi-Agent Systems, December 10–13, 1996, Kyoto, Japan* (pp. 338–344). Menlo Park, CA: AAAI Press.
- Steels, Luc. (1999). *The Talking Heads Experiment*. Volume 1: *Words and meaning*. Special pre-edition for LABORATORIUM, Antwerpen, 1999. Brussels: Artificial Intelligence Laboratory VUB.
- Strawson, Peter. (1950). On referring. *Mind* 59 (235): 320–344.
- Szabolcsi, Anna & Bill Haddican. (2004). Conjunction meets negation: a study in cross-linguistic variation. *Journal of Semantics* 21 (3): 219–249.
- Tomasello, Michael. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Vogt, Paul. (2000). Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication* 4 (1): 89–118.
- Vogt, Paul. (2004). Minimum cost and the emergence of the Zipf-Mandelbrot law. In: Jordan Pollack, Mark Bedau, Phil Husbands, Takashi Ikegami & Richard Watson (eds.), *Artificial Life IX: proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems* (pp. 214–219). Cambridge, MA: MIT Press.
- Vogt, Paul. (2005). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence* 167 (1–2): 206–242.
- Wray, Alison. (1998). Protolanguage as a holistic system for social interaction. *Language & Communication* 18 (1): 47–67.
- Zipf, George Kingsley. (1935). *The psycho-biology of language*. Cambridge, MA: MIT Press.
- Zipf, George Kingsley. (1949). *Human behaviour and the principle of least effort*. Cambridge, MA: Addison-Wesley.
- Zwarts, Frans. (1986). *Categoriale grammatica en algebraïsche semantiek*. Ph.D. dissertation, Rijksuniversiteit Groningen.
- Zwicky, Arnold. (1973). Linguistics as chemistry: the substance theory of semantic primes. In: Stephen Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle* (pp. 467–485). New York: Holt, Rinehart and Winston.
- Żyliński, Eustachy. (1925). Some remarks concerning the theory of deduction. *Fundamenta Mathematicae* 7: 203–209.

Appendices

Appendix 1. The 16 two-placed logical connectives

One DNF-disjunct; Actual situations

<i>symbol</i>	<i>truth function</i>	<i>alt. symbols / names</i>	<i>remarks</i>
\wedge	{11}	&; AND; conjunction	
BUT ⁴⁷	{01}		non-commutative
ALTHOUGH	{01}		non-commutative
\downarrow	{00}	NOR; Peirce's dagger	referentially complete non-confessional

Two DNF-disjuncts; Two belief states

<i>symbol</i>	<i>truth function</i>	<i>alt. symbols / names</i>	<i>remarks</i>
AM	{11,01}	AM = And Maybe	non-commutative $P \text{ AM } Q \equiv P$
MA	{11,01}	MA = Maybe And	non-commutative non-compositional $P \text{ MA } Q \equiv Q$
\leftrightarrow	{11,00}	IFF; XNOR; XAND; biconditional	non-confessional
$\underline{\vee}$	{01,01}	W; XOR; exclusive or	
MN	{01,00}	MN = Maybe Not	non-commutative non-compositional non-confessional $P \text{ MN } Q \equiv \neg Q$
NM	{01,00}	NM = Not Maybe	non-commutative non-compositional non-confessional $P \text{ NM } Q \equiv \neg P$

⁴⁷ I use the symbols BUT and ALTHOUGH for these connectives following Sierra-Santibañez (2001). The words are picked by analogy to the *syntactic* behaviour of *but* and *although* in English: *but* usually allows a positive statement in the antecedent and a negative statement in the consequent, while *although* has the reverse properties. However, the truth-functional meaning of *but* and *although* in natural language equals that of \wedge .

Three DNF-disjuncts; Three belief states

<i>symbol</i>	<i>truth function</i>	<i>alt. symbols / names</i>	<i>remarks</i>
\vee	$\{11,01,01\}$	OR; inclusive or; disjunction	
\rightarrow	$\{11,01,00\}$	material implication; conditional	non-commutative non-confessional
\leftarrow	$\{11,01,00\}$	backward implication	non-commutative non-confessional
$ $	$\{01,01,00\}$	$\underline{\wedge}$; NAND; Sheffer's stroke	referentially complete non-confessional

Four DNF-disjuncts; Four belief states

<i>symbol</i>	<i>truth function</i>	<i>alt. symbols / names</i>	<i>remarks</i>
ALWAYS	$\{11,01,01,00\}$	universal truth	non-compositional non-confessional

Zero DNF-disjuncts; Zero belief states

<i>symbol</i>	<i>truth function</i>	<i>alt. symbols / names</i>	<i>remarks</i>
NEVER	$\{\}$	universal falsehood	non-compositional

Appendix 2. English language lexicalisations of all 16 connectives

Assume the following **basic propositions**:

P	<i>the tiger is coming</i>
Q	<i>I brought a spear</i>

Their **negations** are:

$\neg P$	<i>the tiger is not coming</i>
$\neg Q$	<i>I did not bring a spear</i>

Lexicalisations of connectives representing **actual situations (one DNF-disjunct)**:

$P \wedge Q$	<i>the tiger is coming AND I brought a spear</i>
$P \text{ BUT } Q$	<ul style="list-style-type: none">– <i>the tiger is coming AND I did NOT bring a spear</i>– <i>the tiger is coming but I did NOT bring a spear</i>
$P \text{ ALTHOUGH } Q$	<ul style="list-style-type: none">– <i>the tiger is NOT coming AND I brought a spear</i>– <i>the tiger is NOT coming although I have a spear</i>
$P \downarrow Q$	<i>NEITHER is the tiger coming NOR did I bring a spear</i>

Lexicalisations of connectives representing **two belief states (two DNF-disjuncts)**:

$P \text{ AM } Q$	<ul style="list-style-type: none">– <i>the tiger is coming AND I brought a spear OR I did NOT bring a spear</i>– <i>the tiger is coming AND maybe I brought a spear</i>
$P \text{ MA } Q$	<ul style="list-style-type: none">– <i>the tiger is coming OR the tiger is NOT coming AND I brought a spear</i>– <i>maybe the tiger is coming AND I brought a spear</i>
$P \leftrightarrow Q$	<ul style="list-style-type: none">– <i>the tiger is coming AND I brought spear</i>– <i>OR the tiger is NOT coming AND I did NOT bring a spear</i>– <i>I bring a spear IF and only if the tiger is coming</i>– <i>I only bring a spear if the tiger is coming</i>
$P \underline{\vee} Q$	<ul style="list-style-type: none">– <i>the tiger is coming OR I have a spear, but not both</i>– <i>I never have a spear when the tiger is coming,</i> <i>and when I have one it won't come</i>
$P \text{ NM } Q$	<ul style="list-style-type: none">– <i>the tiger is NOT coming AND I brought a spear OR I did NOT bring a spear</i>– <i>the tiger is NOT coming AND maybe I brought a spear</i>
$P \text{ MN } Q$	<ul style="list-style-type: none">– <i>the tiger is coming OR the tiger is NOT coming AND I did NOT bring a spear</i>– <i>maybe the tiger is coming AND I did NOT bring a spear</i>

Lexicalisations of connectives representing **three belief states (three DNF-disjuncts)**:

- $P \vee Q$ *the tiger is coming OR I have a spear*
- $P \leftarrow Q$ *if I have a spear the tiger is coming*
- $P \rightarrow Q$ *if the tiger is coming then I have a spear*
- $P | Q$ – ***it is NOT the case that** the tiger is coming AND I have a spear*
 – *not both the tiger and the lion are coming*

Lexicalisations of connectives representing **four belief states (four DNF-disjuncts)**:

- $P \text{ ALWAYS } Q$ – *the tiger is coming OR the tiger is NOT coming*
 AND I have a spear OR I do NOT have a spear
 – *this statement is always true*

Lexicalisations of connectives representing **zero belief states (zero DNF-disjuncts)**:

- $P \text{ NEVER } Q$ – *the tiger is coming AND the tiger is NOT coming*
 OR I have a spear AND I do NOT have a spear
 – *this statement is always false*

Appendix 3. A pay-off table leading to an ESSS with separate signals for all 16 connectives

Pay-off tables are representations of how useful actions are in certain situations. They are defined by:

- a set of actions A
- a set of situations T
- a utility function $\forall t(t \in T) \forall a(a \in A): U(t,a) = u$

The pay-off table below is defined in such a way that speakers and hearers will evolve an Evolutionary Stable Signalling Strategy (ESSS) in which there are separate messages for pairs of situations conjoined by each connective.

The pay-off values for the situations $P \wedge Q$, $P \text{ BUT } Q$, $P \text{ ALTHOUGH } Q$ and $P \downarrow Q$ have been assigned arbitrarily. The pay-off values of the other connectives have been calculated as averages of $P \wedge Q$, $P \text{ BUT } Q$, $P \text{ ALTHOUGH } Q$ and $P \downarrow Q$, depending on which of these four is part of the DNF-representation of that connective.

Situations	Actions													
	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉	a ₁₀	a ₁₁	a ₁₂	a ₁₃	a ₁₄
$P \wedge Q$	20.0	0.0	0.0	0.0	18.5	17.1	17.0	0.0	0.0	0.0	13.0	12.5	11.5	0.0
$P \text{ BUT } Q$	0.0	20.0	0.0	0.0	18.5	0.0	0.0	15.5	15.0	0.0	13.0	12.5	0.0	10.5
$P \text{ ALTHOUGH } Q$	0.0	0.0	20.0	0.0	0.0	17.1	0.0	15.5	0.0	14.0	13.0	0.0	11.5	10.5
$P \downarrow Q$	0.0	0.0	0.0	20.0	0.0	0.0	17.0	0.0	15.0	14.0	0.0	12.5	11.5	10.5
$P \text{ AM } Q$	10.0	10.0	0.0	0.0	18.5	8.6	8.5	7.8	7.5	0.0	13.0	12.5	5.8	5.3
$P \text{ MA } Q$	10.0	0.0	10.0	0.0	9.3	17.1	8.5	7.8	0.0	7.0	13.0	6.3	11.5	5.3
$P \leftrightarrow Q$	10.0	0.0	0.0	10.0	9.3	8.6	17.0	0.0	7.5	7.0	6.5	12.5	11.5	5.3
$P \vee Q$	0.0	10.0	10.0	0.0	9.3	8.6	0.0	15.5	7.5	7.0	13.0	6.3	5.8	10.5
$P \text{ NM } Q$	0.0	10.0	0.0	10.0	9.3	0.0	8.5	7.8	15.0	7.0	6.5	12.5	5.8	10.5
$P \text{ MN } Q$	0.0	0.0	10.0	10.0	0.0	8.6	8.5	7.8	7.5	14.0	6.5	6.3	11.5	10.5
$P \vee Q$	6.7	6.7	6.7	0.0	12.3	11.4	5.7	10.3	5.0	4.7	13.0	8.3	7.7	7.0
$P \leftarrow Q$	6.7	6.7	0.0	6.7	12.3	5.7	11.3	5.2	10.0	4.7	8.7	12.5	7.7	7.0
$P \rightarrow Q$	6.7	0.0	6.7	6.7	6.2	11.4	11.3	5.2	5.0	9.3	8.7	8.3	11.5	7.0
$P Q$	0.0	6.7	6.7	6.7	6.2	5.7	5.7	10.3	10.0	9.3	8.7	8.3	7.7	10.5

The condition that the pay-off table conforms to that makes it yield such an ESSS is the following:

For $\forall t_n(t_n \in T)$, $\exists a_1(a_1 \in A)$ such that

- (1) for $\forall a_n(a_n \in A \wedge a_n \neq a_1 \rightarrow U(t_1, a_1) > U(t_1, a_n))$
- (2) for $\forall t_n(t_n \in T \wedge t_n \neq t_1)$, $\exists a_2(a_2 \in A \wedge a_2 \neq a_1 \wedge U(t_n, a_2) > U(t_n, a_1))$

This means that every situation t_1 must have an action a_1 associated with it that has the highest pay-off for t_1 (condition 1), and a_1 must not be the action with highest pay-off for any other situation t_n (condition 2). For each situation t_1 , the highest pay-off for that situation has been set **bold** in the table. The reader may verify that this action that yields this highest pay-off for t_1 is not the action with the highest pay-off for any other situation. For further discussion cf. §4.5.

Appendix 4. The effects of N_T , N_B and N_R on N_{disc} for 10 of the 16 connectives

The graphs depict the effects of number of topic objects N_T (horizontal axis of each graph), number of background objects N_B (separate lines in each graph) and number of meaning space regions N_R (separate graphs) on the frequencies N_{disc} of 10 of the 16 possible two-placed connectives. The graphs of \leftarrow , ALTHOUGH, AM/MA and NM/MN are identical to those of \rightarrow , BUT, TRUE and \neg respectively, while ALWAYS and NEVER yield $N_{disc} = 0$ in all situations, as detailed in §7.3.1 and §7.3.4. The connectives are all ranked in order of specificity. This page shows the positive connectives. The next page of the appendix shows the negative connectives. See §7.3.7 and §7.3.5 for more information on the notions of specificity and positive and negative connectives, respectively.

